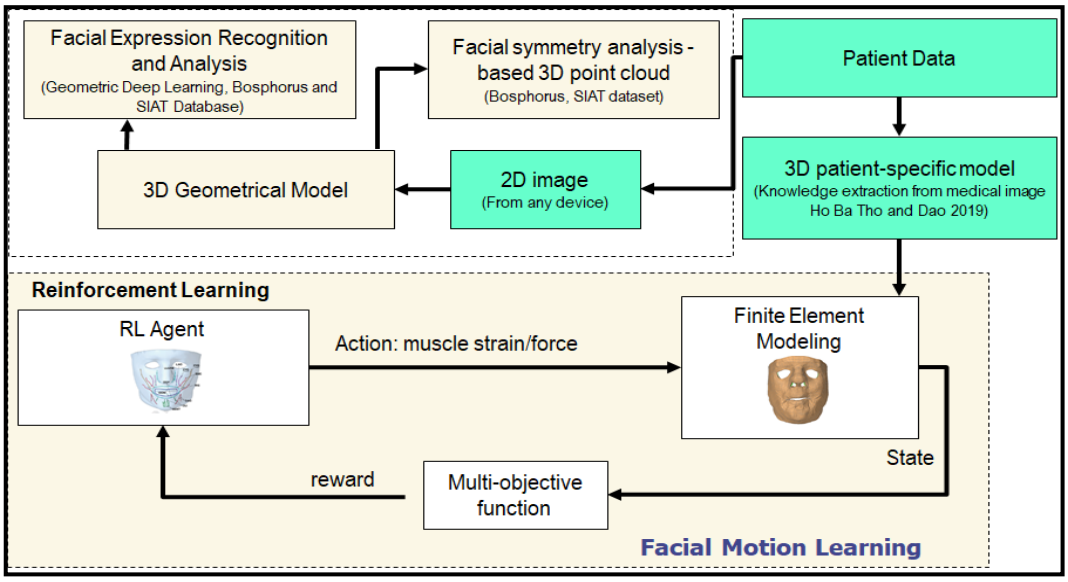
Enhanced facial behavior recognition and rehabilitation using 3D biomechanical features and deep learning approaches

Duc-Phong Nguyen

Par **Duc-Phong NGUYEN**

Enhanced facial behavior recognition and rehabilitation using 3D biomechanical features and deep learning approaches

Thèse présentée
pour l'obtention du grade
de Docteur de l'UTC



Soutenue le 30 juin 2022
Spécialité : Biomécanique et Bioingénierie : Unité de Recherche Biomécanique et Bioingénierie (UMR-7338)
D2688

Enhanced facial behavior recognition and rehabilitation using 3D biomechanical features and deep learning approaches

Thesis defended on 30 June 2022 in Biomechanics and Bioengineering

Spécialité : Biomécanique et Bioingénierie

Doctoral school ED71 of the University of Technology of Compiègne

Duc-Phong NGUYEN

Jury Members

VAN DER SLOTEN Jos, Professeur, Biomécanique, KU Leuven (Rapporteur)

BURDIN Valérie, Professeur, Biomécanique, IMT Atlantique (Rapportrice)

NGUYEN Sao-Mai, Maître de conférences, Robotique et IA, Institut Polytechnique de Paris (Examinatrice)

MOHAMED DAOUDI, Professeur, Signal and Image, IMT Nord Europe (Examinateur)

EL KIRAT Karim, Professeur, Université de technologie de Compiègne (Examinateur)

DAO Tien-Tuan, Professeur, Biomécanique, Ecole Centrale Lille (Directeur de thèse)

HO BA THO Marie-Christine, Professeur, Biomécanique, Université de technologie de Compiègne (Co-directeur de thèse)

DAKPÉ Stéphanie, Professeur, Maxillofacial Surgery, CHU Amiens (Membre invité)

Acknowledgement Integrity

First and foremost, I would like to extremely express my deepest and strongest gratitude to my supervisors, Professor Tien-Tuan Dao and Professor Marie-Christine Ho Ba Tho, for their guidance, enthusiasm, encouragement, dedication, and relentlessly support throughout the course of my Ph.D. program. My thesis could not be completed without their support.

I would like to sincerely thank my committee members, Professor Van Der Sloten Jos, Professor Burdin Valérie, Professor NGUYEN Sao-Mai, Professor Mohamed Daoudi, Professor El Kirat Karim, and Professor Dakpé Stéphanie for their insightful comments and spending their precious times to evaluate my thesis. Your evaluations not only improved my thesis but also suggested my future works.

I also would like to acknowledge the Université de Technologie de Compiègne and the Doctoral School for their administrative support. I also thank the Biomechanics and Bioengineering Laboratory (BMBI -UMR CNRS 7338). Throughout my Ph.D., the Laboratory organized professional working spaces and provided helpful research resources. Additionally, I want to express my appreciation to the CHU Amiens Hospital, in particular to Professor Stéphanie Dakpé and Colin Emilien, for their assistance with collecting the database.

Thanks to Dr. Tan-Nhu Nguyen, Mr. Diego Alfredo Quexada Rodriguez, Ms. Katharine Nowakowski, Dr. Dang-Phong Bach, Dr. Cong-Uy Nguyen, and Dr. Tuan-Anh Bui for always being willing to help me with scientific discussions at the Université de Technologie de Compiègne as well as my life in France. I will always cherish the wonderful times I had with friends and colleagues while pursuing my doctorate.

Moreover, I would like to acknowledge Sorbonne Center for Artificial Intelligence (SCAI) for their financial support to the course of my Ph.D. program.

Last but not least, I am really extremely grateful to my parents, Duc-Huy Nguyen and Thi-Lan Bui, as well as my brother, Duc-Duy Nguyen and his wife Ngoc Tram, Duc-Canh Nguyen and his wife Van-And Nguyen, and my girlfriend Thi Thuy Duong Pham for their unwavering support and patience throughout my life. I shall never forget their devotion for me and will always be proud of them.

Abstract

Facial palsy patients or patients under facial transplantation have facial dysfunctionalities and abnormal facial motion. This is due to the altered facial nerve and facial muscle systems. Current traditional facial rehabilitation is based mainly on a mirror approach to monitor the visual qualitative feedback from the rehabilitation exercise. Computer-aided systems based on physics-based models have also been developed to provide objective and quantitative information. However, the use of these systems in clinical routine practice still remains challenging due to several limitations: 1) the lack of building 3D information from images or the dependency on the selected cameras; 2) the lack of analyzing the face in terms of expression recognition and symmetry analysis and 3) the limitation of the predictive capacity of the facial motion patterns with emerging biomechanical properties.

The objectives of this Ph.D. are to develop innovative engineering solutions toward a next-generation computer-aided decision support system for facial analysis and rehabilitation. The thesis provided four main contributions: 1) the fast reconstruction of the 3D face shape from a single 2D image using deep learning approaches; 2) the improvement of facial expression recognition using 3D point set and geometric deep learning; 3) the face symmetry analysis based on novel descriptors and the difference between two different populations (Caucasian and Asian) and 4) the proposition of a novel modeling workflow for learning facial motion by coupling reinforcement learning and the finite element modeling for facial motion learning and prediction.

The thesis opens new avenues for developing and deploying an end-to-end computer-aided decision support system to guide and optimize the functional rehabilitation program.

Keywords: 3D face shape reconstruction, 3D facial expression recognition, 3D facial symmetry analysis, facial mimic rehabilitation, geometric deep learning, reinforcement learning.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

2D	Two-Dimension
3D	Three-Dimension
4E	4 basis expressions
6E	6 basis expressions
3DMM	3D Morphable Model
AAM	Active appearance model
Acc	Accuracy
Acc-val	Accuracy on validation dataset
AI	Artificial Intelligence
AFM	Annotated Face Model
ANNs	Artificial Neural Networks
ANs	Actor Networks
ASM	Active appearance model
AUs	Action Units
BBN	Bayesian belief network
BEs	Basic expressions
BFM	Basel Face Model
BP4D	Binghamton-Pittsburgh 3D dynamic spontaneous facial expression
BU-3DFE	Binghamton university 3D facial expression
BU-4DFE	Binghamton university 3D dynamic facial expression
CDF	Cumulative distribution function
CDSSs	Computer-aided Decision-Support Systems
CNNs	Convolutional Neural Networks
CEs	Compound expressions
CK	Cohn Kanade
CK+	Cohn Kanade Extension
CNs	Critic Networks
CPU	Central Processing Unit
CT	Computed Tomography
CUDA	Compute Unified Device Architecture
DBN	Deep belief network
DDPG	Deep deterministic policy gradient
DECA	Detailed Expression Capture and Animation
DL	Deep learning
DMCMs	Differential Mean Curvature Maps
Doi	Digital Object Identifier
DRML	Deep region and multi-label learning
eFace	Clinician-graded electronic Facial Paralysis
FacCE	Facial Clinimetric Evaluation
FACS	Facial Action Coding System
FC	Fully connected layer
FDI	Facial Disability Index
FE	Finite Element
FEM	Finite Element Method
FER	Facial expression recognition
FLAME	Faces Learned with an Articulated Model and Expressions
FLs	Facial landmarks
FML	Facial motion learning

FSA	Facial symmetry analysis
IF	Impact factor
GAN	Generative Adversarial Network
GPU	Graphics Processing Unit
HD	High Definition
HMMs	Hidden Markov Models
HOG	Histogram of oriented gradient
KdTree	K-dimensional Tree
LBP	Local binary pattern
LDA	Linear Discriminant Analysis
LSFM	Large Scale Facial Model
L/R BUC	Left/Right Buccinator
L/R DAO	Left/Right Depressor Anguli Oris
L/R DLI	Left/Right Depressor Labii Inferioris
L/R LAO	Left/Right Levator Anguli Oris
L/R LLSAN	Left/Right Levator Labii Superioris Alaeque Nasi
L/R MENT	Left/Right Mentalis
L/R OOM	Left/Right Orbicularis Oris Marginalis
L/R OOP	Left/Right Orbicularis Oris Peripheralis
L/R ZYG	Left/Right Zygomaticus
L/R RIS	Left/Right Risorius
LSTM	Long-short term memory
LYHM	Liverpool-York Head Model
MEs	Micro-expressions
MLP	Multi-layer perceptron
MRI	Magnetic Resonance Imaging
PC	Principal Component
PCA	Principal Component Analysis
PDF	Probability density function
RAM	Random Access Memory
RBF	Radial Basic Functions
ResNet	Residual neural network
RGB	Red Green Blue
RGB-D	Red Green Blue-Depth
RL	Reinforcement Learning
RNNs	Recurrent Neural Networks
SA	Set abstraction
SFAM	Statistical facial feature models
SFGS	Sunnybrook Facial Grading System
SFM	Surrey Face Model
SIFT	Scale-invariant Fourier transform
SOP	Scaled orthographic projection
STL	Standard Triangle Language
SVM	Support Vector Machine
TD3	Twin delayed DDPG
VGG-Face Network	Visual Geometry Group-Face Network

Chapter 1:

Introduction

The aim of the Ph.D. project is to develop a clinical decision support system for rehabilitation of facial palsy combining artificial intelligence and biomechanical knowledge.

This project contributes to one of sustainable development goals. In particular, goal 3 **good health and well-being** which aims to improve the health quality and promote well-being for all at all ages.

The first chapter discusses the clinical circumstances of facial palsy and facial transplantation patients. The chapter also discusses practical limitations of traditional face rehabilitation methods as well as modern computer-aided systems. Machine learning, deep learning, and reinforcement learning were also used for clinical applications. The goal of the project is to provide a computer-aided decision support system for facial mimic rehabilitation based on the use of artificial intelligence.

1.1. Clinical context and needs

1.1.1. Facial palsy, facial transplantation, and facial mimic rehabilitation

Communication is essential to human development and personal life [1]. Effective communication can also help strengthen personal and professional connections [2]. Nonverbal components (such as facial expression and body language) account for 55% of interpreting human communication, whereas the tone of voice and words account for 38% and 7%, respectively, according to the *7%-38%-55% communication rule* [3]–[8]. Furthermore, facial expressions have an important role in a person's identity, communication, and health [9]. In fact, patients with facial palsy or those who have undergone facial transplant surgery have abnormal facial motion patterns. This is because of altered facial muscle functions and nerve damage [10]–[13]. As a result, these involved patients have had their personal, professional, and social lives negatively impacted [10]–[13]. Facial expressions, in particular, are facial skin deformations. They are results of facial muscle contractions and/or relaxations spanning between insertion points on the soft tissues (skin layers) and attachment points on the bone (cranial layers) [14]–[16]. Electrical signals from the seventh cranial nerve control the contraction of these muscles [17]. Faulty nerves can cause several phenomena. The first one is atypical facial functions such as undesirable facial motions during eating, speaking, and expressing oneself. The second one is unnatural relaxations of the cheeks, mouth angle, and eyelids [18]. The third one is asymmetrical mimics when kissing, smiling, and pronouncing words [19]. This leads to abnormal facial movements, as well as psychological implications and a loss of quality of life for involved patients.

Figure 1. Facial palsy patient (left) with dropping mouth corner and facial transplantation patient (right) [20].

Each year, approximately 20 to 30 people per 100,000 are affected by facial nerve paralysis [21]. About 1.5 percent of the population is afflicted at some point in their lives [13]. The involved patients are typically between the ages of 31 and 60, with some pregnant

women included [22]. The majority of people with facial palsy (80-84 percent) have modest symptoms that can be resolved on their own or with medicine within a few weeks or months. But some (16-20 percent) have more severe symptoms that are permanent [23].

There are many possible factors that hinder the facial nerve from working properly in facial palsy patients. These include infections, viruses, trauma, post-surgery, and any reason that produces pressure on the facial nerve[24]. This disrupts the nerve delivering signals from the brain to the facial muscles. Patients who have had a part or all of their face transplanted are more likely to develop facial paralysis [20], [25]–[32]. On November 27, 2005, Prof. Bernard Devauchelle of the CHU Amiens-Picardie in France performed the first-ever facial transplantation (Figure 2a) [20]. The patient has a good sensory function with cold and hot after 6 months of surgery. After 10 months, she can close her mouth and grin after 18 months. The patient was pleased with the aesthetic, sensory and motor functional, as well as psychological outcomes after 18 months of transplantation. Dr. Maria Siemionow of the Cleveland Clinic in the United States completed the first near-total facial transplantation in the United States in December 2008. The surgery replaced an 80 percent facial loss in a 22-hour operation (Figure 2b) [25]. After 8 months of post-transplantation care, both physical and psychological conditions improved dramatically. The patient had good nose smelling, mouth tasting, speaking, drinking, and eating functioning. In April 2018, Dr. Laurent Lantieri of the Georges Pompidou Hospital in France conducted second face transplantation for one patient (Figure 2c) [26]. The first time was in 2010 when the patient was 35 years old with the donor whose age was 60. The second time was in 2018 after the injection of the first transplanted face with the donor whose age was 22. In ten years since Prof. Bernard Devauchelle performed the first facial transplantation in France, a total of 37 cases in the wide world has been performed, with 20 partial and 17 full face transplantations [27], [28]. The patients had a variety of injuries prior to transplantation. These include ballistic trauma, burns, vascular lesions, animal attacks, neurofibromatosis, traumatic face trauma, and cancer/radiation therapy sequelae. Sensory capabilities such as hot, cold, painful sensations, and light touch can be restored in 3 to 8 months after facial transplantation. However, motor functions like feeding, swallowing, speaking effectively, smiling, and expressions take longer, ranging from 2 to 8 years. It is important to emphasize that an appropriate rehabilitation program for improving face functions in relation to motor functions should be devised.

A **face rehabilitation program** must be individually constructed in terms of both physical and psychological clinical stages [19], [33]. Individual assessment of the severity of facial paralysis degrees using **clinical** or **non-clinical grading systems** is the first stage in facial rehabilitation. For *clinical facial grading*, various facial grading methods have been proposed. Sunnybrook Facial Grading System, for example, ranges the score from 0 (absolute face paralysis patients) to 100 (normal patients) [34]. The system assesses facial symmetry, as well as synkinesis at rest position and during voluntary movements. The House-Brackmann system is divided into six categories, ranging from I (normal patients) to VI (total paralyzed patients), [35], [36]. The evaluation is measured in both static and dynamic expressions. Other grading systems were based on questionnaires, such as Facial Clinimetric

Evaluation (FacCE) [37] and Facial Disability Index (FDI) [38]. Patients complete the self-report for the assessment of facial nerve disabilities. Computer-based systems are also used for *non-clinical facial grading systems* from data collected by electronic sensors [39]. Individual treatment plans will be created based on the patient's facial paralysis exam and will include five primary therapy components [33]. 1) The first component is *patient education*. This teaches patients about facial anatomy, including the nerves and muscles of the face, and how they work together during facial movements and expressions. 2) The second component is *soft tissue mobilization*, also known as massage treatment. This helps to relax muscles, promote healing, and break down scar tissue, hence enhancing mobility, comfort, and reducing stress in the afflicted face muscles. 3) The third component is *functional retrain*. Patients are taught fundamental mouth motions such as lip closure, the capacity to eat, drink, and communicate. 4) The fourth component is *facial expression retrain*. Patients were instructed to produce initial facial expressions such as smiles and sound pronunciations. Finally, 5) *synkinesis management*, in which patients were touched to regulate undesirable facial movements.

Figure 2. Several examples of patients before and after having facial transplantation: (a) Prof. B. Devauchelle (CHU Amiens, France) conducted the first-ever facial transplantation in 2005 [20]. (b) Dr. Maria Siemionow (Cleveland Clinic, USA) conducted the first case of facial transplantation in the United States in 2008 [25]. (c) Dr. Laurent Lantieri (Georges Pompidou Hospital Paris, France) conducted two transplantations for a male patient (the first surgery was in 2010 with the face of the donor was 60 and the patient was 35, and the second surgery was in 2018 with the face of the donor was 22) [26].

Despite many studies, facial mimic rehabilitation continues to be challenging at all stages. *In the stage of facial paralysis grading*, the paralysis assessment based on the traditional grading systems is accessible. This is simple to use and demands low-cost equipment. However, this procedure is inherently inexact and subjective. Furthermore, non-clinical facial grading systems that use objective computer-aided systems have mainly used 2D images to examine the face [39]. A computer-based system that detects action units (AUs) described by the Facial Action Coding System (FACS) [40] can be used to assess patients. The recognition of AUs, on the other hand, was dependent on image processing. Therefore, facial behavior recognition was performed only with static and dynamic image databases. This approach necessitates feature engineering, but no direct processing on 3D databases. *In the stage of treatment*, the traditional rehabilitation is a long-term, inconvenient for the patient, and

occasionally inefficient procedure throughout the therapy stage. Most therapies include face-to-face encounters between patients and their therapists in order to practice facial mimics. Patients get demotivated as a result of these recurrent training sessions for manipulating specific face muscles [19]. It's worth noting that most facial rehabilitation has mainly been based on a mirror approach to monitor the visual qualitative feedback from the rehabilitation activity. Patients utilize their distorted features in the mirror as a reference to teach themselves the correct expressions, making it difficult to practice appropriately. Furthermore, traditional rehabilitation is also particularly challenging owing to the lack of quantitative and objective tools that provide bio-information such as facial muscle excitation and contraction.

1.1.2. Development of machine learning, deep learning, and reinforcement learning and clinical applications

This section covers fundamental information about artificial intelligence, machine learning, deep learning, reinforcement learning, and the application of these approaches to real-world problems, particularly in healthcare settings.

1.1.2.a. Artificial intelligence, machine learning, and deep learning

John McCarthy was the first to coin the term artificial intelligence (AI) in 1955. The engineering and science of creating intelligent devices, particularly intelligent computer programs, is referred to as AI [41]. This is the capacity of a machine with an intelligent conception to learn, plan, and reason, as well as sense or perceives knowledge. AI aims to automate intellectual tasks that are typically performed by humans. Machine learning and deep learning are subfields that help AI attain this objective [42] (Figure 3). Conventional programming was explicitly programmed by humans using existing knowledge (Figure 4a). In contrast, machine learning trains input data to generate algorithms that can automatically discover patterns in the input data (Figure 4b).

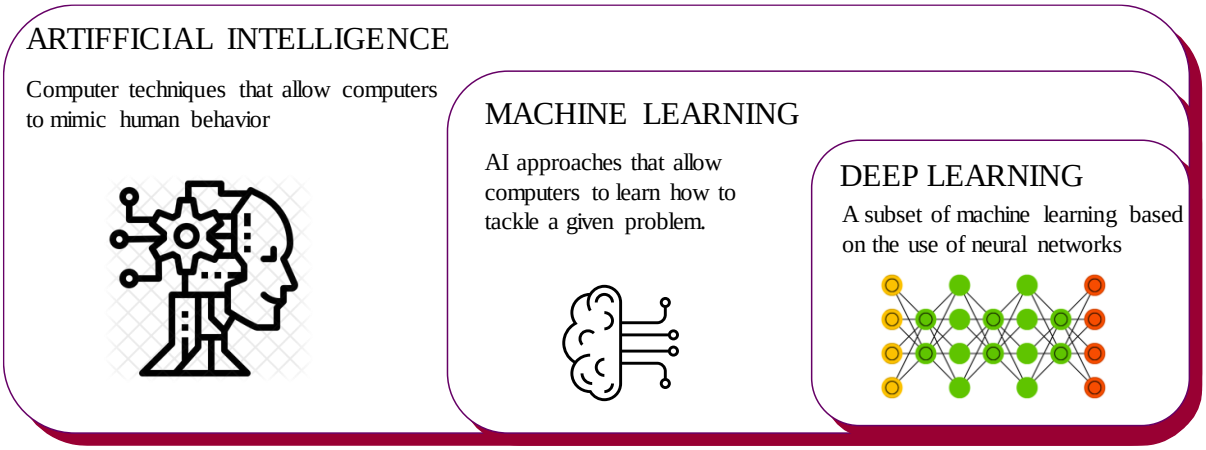


Figure 3. Artificial intelligence, machine learning, and deep learning

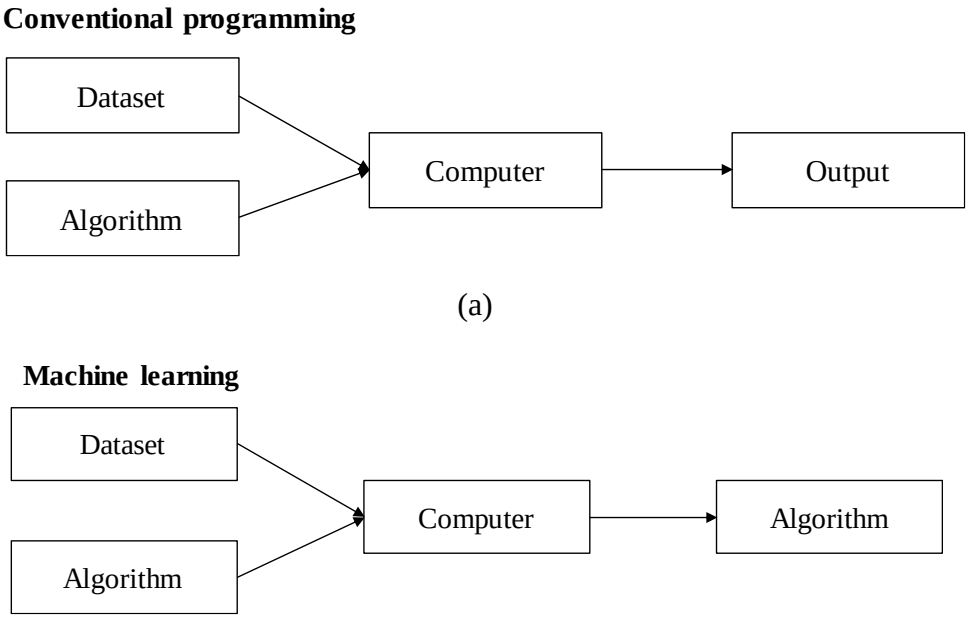
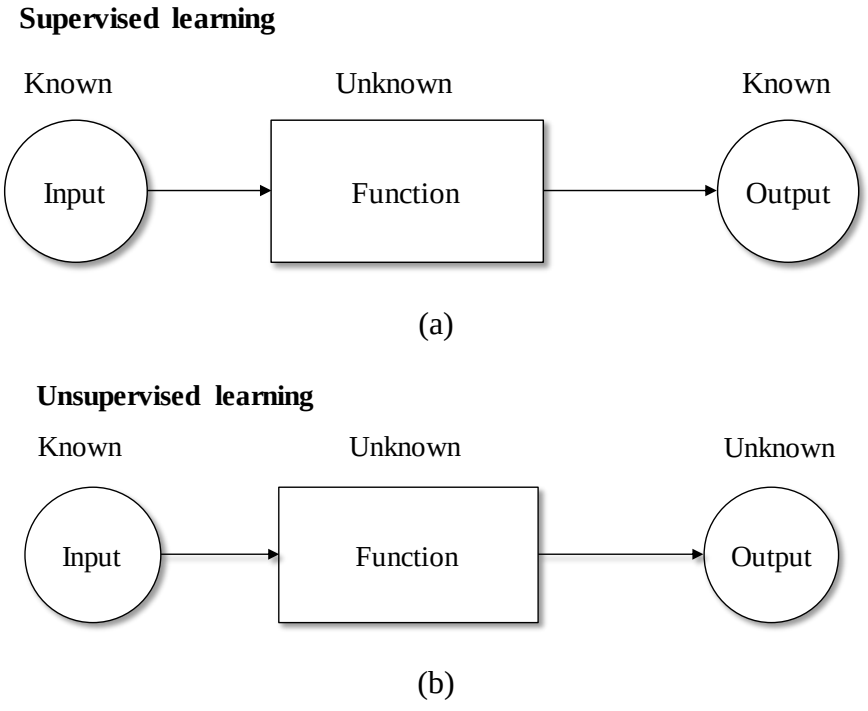
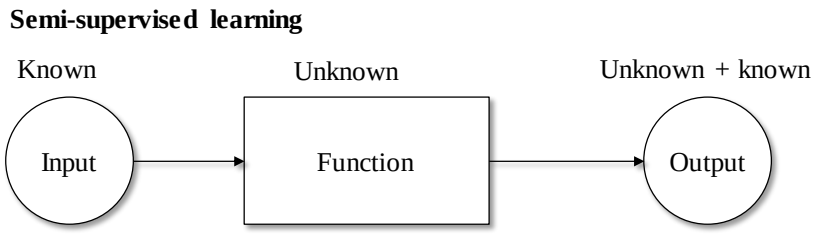


Figure 4. The basic principle of conventional programming (a) and machine learning algorithms (b). Machine learning learns the algorithm that best presents the input data, while conventional programming is explicitly programmed by humans.

There are 4 different types of machine learning (Figure 5), depending on how to learn the data: supervised learning, unsupervised learning, semi-supervised learning, and reinforcement learning *[43]*.





(c)

Figure 5. The basic principle of different types of machine learning algorithms includes supervised learning (a), unsupervised learning (b), and semi-supervised learning (c).

Supervised learning is a sort of machine learning in which the data is fed into the training model together with the matching label (Figure 5a). This helps to learn the mapping function from the input data to the output label. Mathematically, from a set of input $\boldsymbol{X}$ and corresponding output $\boldsymbol{Y}$, supervised learning learns to identify the best mapping function f: $\overline{\boldsymbol{Y}} = f(\boldsymbol{X})$ such that $\overline{\boldsymbol{Y}} \approx \boldsymbol{Y}$. Supervised learning is usually used for classification and regression problems.

Unsupervised learning, on the other hand, is a sort of machine learning in which the model only feeds the input data during training without corresponding output data (Figure 5b). Unsupervised learning algorithms learn the distribution and detect the similarity among input samples. It then separates them into different groups based on the distribution and similarity index. Unsupervised learning algorithms are frequently applied for clustering and dimensionality reduction.

Semi-supervised learning is another type of machine learning, which trains the model that feeds the input data only just a portion of the data being labeled (Figure 5c). Semi-supervised learning is a combination of partially unsupervised learning and partially supervised learning.

Finally, **reinforcement learning**, unlike the other frameworks that deal with static datasets, is unique in that it operates in dynamic situations. Instead of clustering data or finding the label of data, reinforcement learning seeks to determine the greatest feasible sequence of actions that provide the best results.

On the other hand, **deep learning** is a sub-domain of machine learning that used to be trained on a large dataset to perform predictions [44] (Figure 3). Deep learning aims to model the data using complex architectures inspired by artificial neural networks, which are much like the human brain [45].

1.1.2.b. Reinforcement learning

Reinforcement learning is a particular type of machine learning model that makes a sequence of decisions. Reinforcement learning, in particular, is learning strategies to map the evaluations and observations from the environment to a set of actions. This maximizes long-term cumulative rewards (Figure 6). The method employs agents to interact with an action

from a set of actions in an environment resulting in changes in the agent's states. The agents learn by maximizing the cumulative reward obtained from the environment.

Collecting the greatest amount of reward necessitates the agent must explore by interacting and learning from the environment. Initially, the agent produces an action that impacts the environment and alters its state. Afterward, the environment feedbacks by rewarding the action. Utilizing this reward assists the agent adjusts actions to produce in the future. This approach can help the agent acquire experience. It's clear that the agent has a function that receives state observations as input and converts them to actions as outputs. The *policy* is the name given to this mapping function. The policy selects which action to take based on a series of observations. In the beginning, the agent has a poor vision of the environment, the policy may at first be a function that incorrectly maps the state to the action. After that, using reinforcement learning methods, the agent is trained to update in order to achieve the best function for the policy. The agent's actions, the associated observations from the environment, and the value of rewards gathered are used to update the policy. After being updated, the policy is finally able to determine the most effective action from any given condition in order to maximize long-term reward.

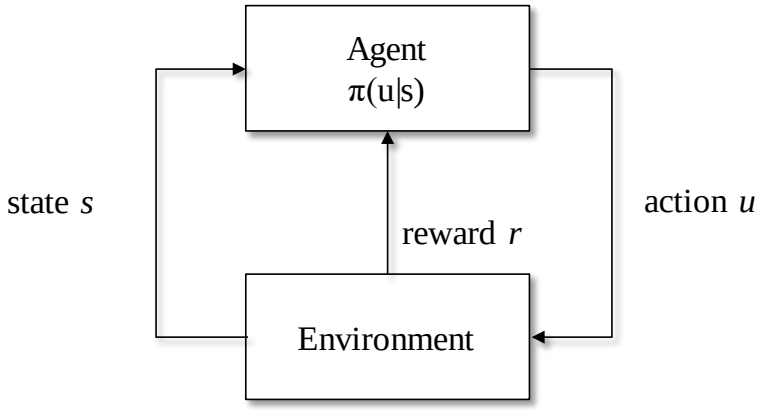


Figure 6. Main components of reinforcement learning include agent, environment, a set of actions, a set of state, and the reward

The reinforcement learning process consists of several components. Firstly, an **environment** is a place where the agent can learn through interaction. Therefore, either simulation or a real physical installation must be used to create the environment's properties. Secondly, it requires a clear **set of actions** that the agent can conduct as well as collect the corresponding **reward**. Thirdly, the **policy** should be determined to structure the parameters and the **training algorithm** to persuade the optimal policy. Finally, the agent implements the policy to **exploit** the environment and attain the desired result.

To begin with, the **environment** may be defined as anything outside of an agent (Figure 7). Through simulation or a model of the world, an agent learns by interacting with it. The learning process necessitates a large number of trial-error-correction trials, maybe in the millions or even tens of millions. By using a simplified environment model and running in

parallel, the simulation environment can run considerably quicker than in the real world. Furthermore, it may imitate situations that are difficult to test in the actual world.

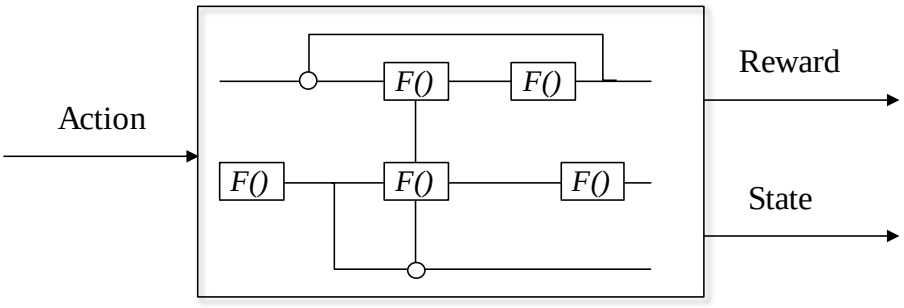


Figure 7. Environment can be modelized by function, simulation

Second, after establishing the environment, **a set of actions** should be devised, as well as the reward derived from the environment. This may be accomplished by implementing a reward function. The agent can receive a sparse reward, instantaneous reward after every time step, or rewards after at end of the episode. The reward is the instantaneous benefit of the agent at a specific observation state. Besides, **value** is the total reward that the agent expects to collect from being in that state in the future. The optimization method tends to assist the agent in choosing the action to assess the value rather than the immediate reward. This allows the agent to collect the most reward over time rather than in a short period of time.

The **policy** and the learning algorithm make up the agent (Figure 8). The policy is a function that takes observations or states as inputs and produces actions as outputs. A simple table termed Q-function, which displays the value of the observation-action pair is one approach to depict this policy function. The agent selects the most valuable state from which to take action. However, if the action-state combination becomes extremely huge, even infinite, as it evolves over time, this type of training may fail. When the state-action gaps are big, building the Q-table is impracticable. This is known as the "curse of dimensionality." It is feasible to use Deep Neural Networks to create a generic function approximator.

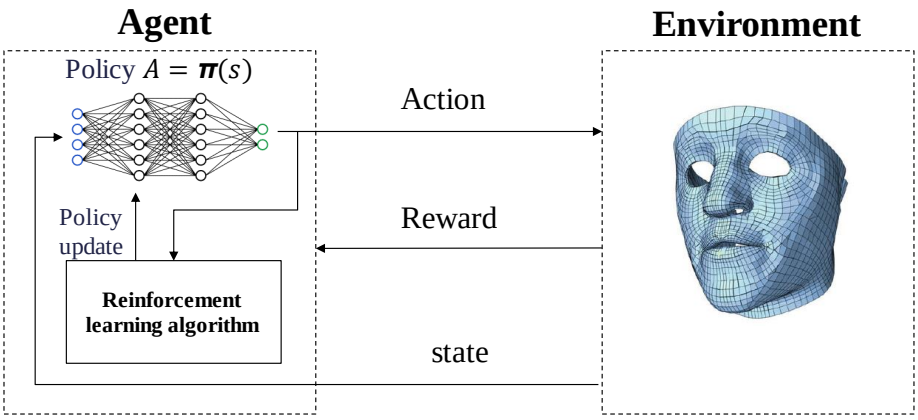


Figure 8. A general architecture of reinforcement learning algorithm.

A neural network is made up of a set of nerve or artificial neurons that link as a universal function approximator to simulate the way a human brain works (Figure 9). A network may be able to map an input into an appropriate output by combining the proper neuron and connection types.

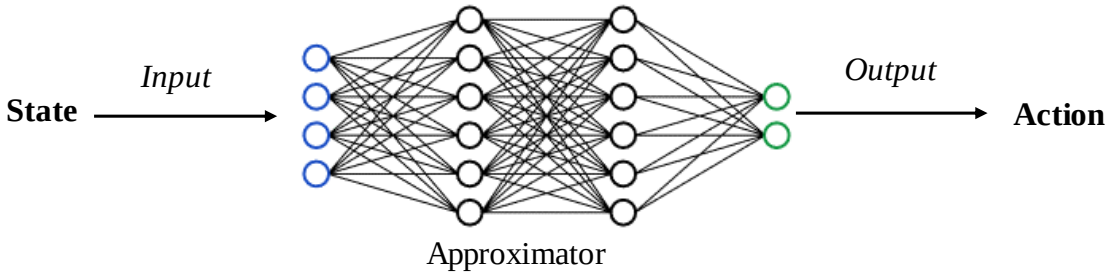


Figure 9. The policy as an approximator takes state as input and outputs an action.

The network is similar to a function that receives a huge number of observation states and converts them to a set of actions. The neural network would learn over time because of the nature of this function. A kind of activation function, the number of neurons in each layer, and the number of hidden layers should all be defined in the network's structure. As a result, the problem should be solved using an appropriate neural structure.

1.1.2.c. Apply of machine learning, deep learning, reinforcement learning in general and in clinical applications

Machine learning, deep learning, and reinforcement learning have been widely used in computer vision, marketing and advertising, natural language processing, finance, transportation, education, and, most notably, healthcare, as huge data and computing capability have advanced.

Computer vision concerns high-level picture and video comprehension [46]. Many applications in computer vision are solved by deep learning approaches, such as Google Brain for image recognition, image qualitative and quantitative enhancement [47]; Facebook's DeepFace [48], and Google's FaceNet [49] for face recognition for commercial reasons.

Commercial businesses use big data and deep learning to improve e-commerce systems, making it simpler for users to search for products and propose suitable things [50], [51].

Natural language processing is being developed with applications in mind, such as machine translation [52], [53], and speed recognition for automated captioning on YouTube or voice assistants (e.g. Siri, Alexa, Cortana).

In the *financial ecosystem*, machine learning, deep learning, and reinforcement learning play important roles in stock price prediction [54], [55], fraud detection for transactions, trading, and credit cards [56], [57].

For *transportation*, companies such as Tesla, Waymo, Argo AI, and others are conducting extensive research and development to develop robust assistance systems. We can mention several dependable services such as traffic accident prediction [58], traffic congestion [59], pedestrian and obstacle detection [60], and traffic sign detection [61].

Deep learning has recently demonstrated expert diagnosis levels in physical conditions. It also shows the ability to construct algorithms that target specific treatments to advise doctors and patients. With the intensive development of wearable medical sensors or record devices, cutting-edge deep learning algorithms can work directly on raw datasets to automatically learn and extract features. Those algorithms include convolution neural networks, recurrent neural networks, long-short term memory, radial basis function, generative adversarial network, etc. This can potentially open a new avenue for real-time diagnostic and rehabilitation interventions [62], [63]. Machine learning and deep learning can diagnose disorders including diabetic retinopathy [64], pneumonia [65], and skin cancer [66] using medical imaging analysis. The development can go even further. The first example is classifying patients regarding the endotype based on modeling molecular data [63]. The second example is defining and analyzing the ontology and mechanism of diseases [67]. A number of clinical practices have already emerged. InnerEye of Microsoft provides a system with a graphical user interface that supports radiologists to assess and diagnose cancerous tumors and arrange surgical interventions [68]. The cooperation between DeepMind Healthy and Moorfields Eye Hospital runs models for retinal pathological diagnosis using optical coherence tomography scanners [69]. IBM's Watson [70] keeps track of health conditions in order to provide individualized cancer care. Reinforcement learning, in particular, is being used in healthcare to deal with medical diagnosis, dynamic treatment regimes, and decision-making tasks [71]. Sequences of rules that monitor healthcare decisions such as treatment strategy and medicinal doses are included in dynamic treatment regimes [72]. Based on clinical observations and patient evaluations, reinforcement learning has helped to build automated decision-making treatment regimes for various chronic illnesses, including HIV [73], diabetes [74], and cancer [75]. Most machine learning algorithms require a huge quantity of data for training. Reinforcement learning, however, aids medical diagnostics by using agents to interact with the environment to create labeled data. Reinforcement learning also aids in the development of decision-making systems for medical diagnosis. These systems were based on medical imaging for image segmentation [76], natural language processing for clinical text data for diagnosing inferences [77], human-computer interface for dialogue systems and chatbots [78], and personalized health recommendation systems for consultation, dosage, and healthy activities [79]. In general, **the integration of machine learning, deep learning, and reinforcement learning in clinical settings and healthcare have improved the safety and efficiency of existing clinical and healthcare systems**.

1.1.2.d. Clinical computer-aided decision support systems

A clinical computer-aided decision support system (CDSS) aims to improve healthcare delivery by reinforcing medical or rehabilitation therapy based on clinical knowledge, patient-specific diagnoses, and other health references [80]. In particular, many computer-

aided decision support systems were developed as **serious games** that could be used for both patient and non-patient-oriented (Figure 10) [81].

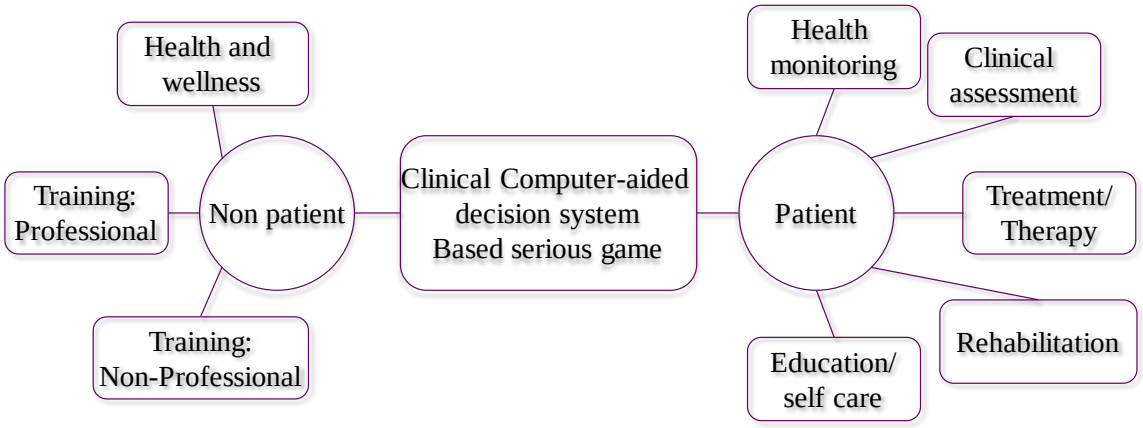


Figure 10. Clinical computer-aided decision support system-based serious games classification [81].

Computer-aided systems are applied for **non-patient** can be used for 3 different categories:

1) *Health and wellness system devices* contain several specific functions that monitor the health of the users such as exercise, sufficient sleep, body weight, alcohol consumption, and smocking. Several examples could be illustrated such as Sensory gate-ball game [82], Fitness adventure [83], Dancing in the streets [84], Virku – Virtual fitness Center User Interface [85].

2) *Training and simulation for professional system devices* are of help doctors or nurses with learning and practicing tools such as Virtual dental implant training simulation program [86], and HumanSim [87].

3) *Training and simulation for non-professional system devices* are of help for raising awareness of players about healthcare such as Hand hygiene training [88], and Nutri-trainer [89].

Computer-aided system-based serious games are applied to **patients** and can be used for 5 different categories:

1) *Health monitoring system devices* are tracking bio-signals of patients for monitoring their health conditions. For example, we can mention software applications such as Heart failure tele-management system [90], and healthcare monitoring [91].

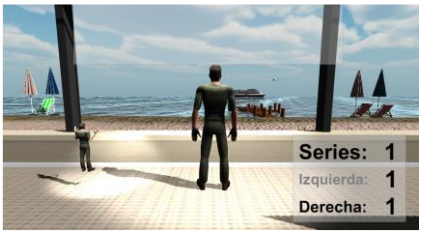
2) *Clinical assessment or detection system devices* can diagnose the patient's symptoms of irregularity such as PlayWithEyes [92], Unobtrusive health [93], and EEG-based serious games [94].

3) *Treatment or therapy system devices* are using for health problems such as diagnosis and management of Parkinson [95], and children's speech disorder therapy [96].

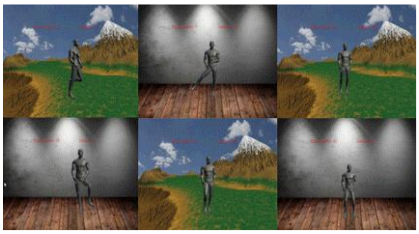
4) *Rehabilitation system devices* assist in restoring functional health and skills after illness such as virtual reality rehabilitation [97], lower limb rehabilitation [98], upper limb

rehabilitation [99], home-based physical exercises [100], [101], and cerebral palsy rehabilitation [102] (Figure 11).

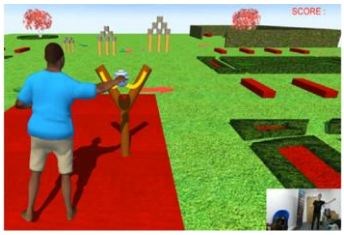
5) *Education or self-care system devices* assists to have a better knowledge of the sickness or health condition, as well as learning how to get and remain healthier while dealing with it. Several systems are namely Serious game for diabetes [103], and cognitive rehabilitation exercises [104].



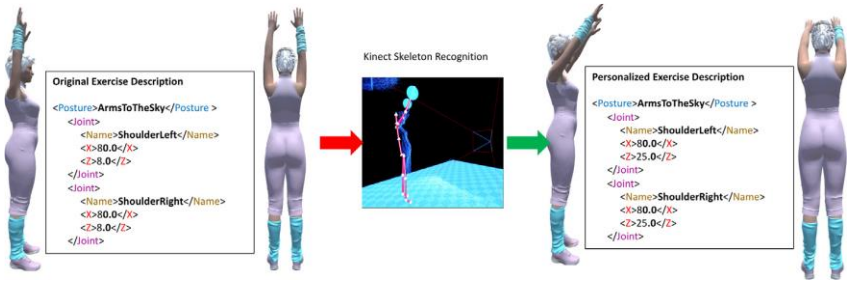
Virtual reality rehabilitation [Pedraza-Hueso 2015]



Lower limb rehabilitation [Tannous 2016]



Upper limb rehabilitation [Esfahlani 2018]



Home-based physical exercises [Fernandez 2018]

Figure 11. Rehabilitation-oriented decision support systems

In new research, Nguyen 2020 proposed a clinical decision support system that can be used for facial mimic rehabilitation which offers cheap cost, easy-to-use, and portable requirements (Figure 12) [105].

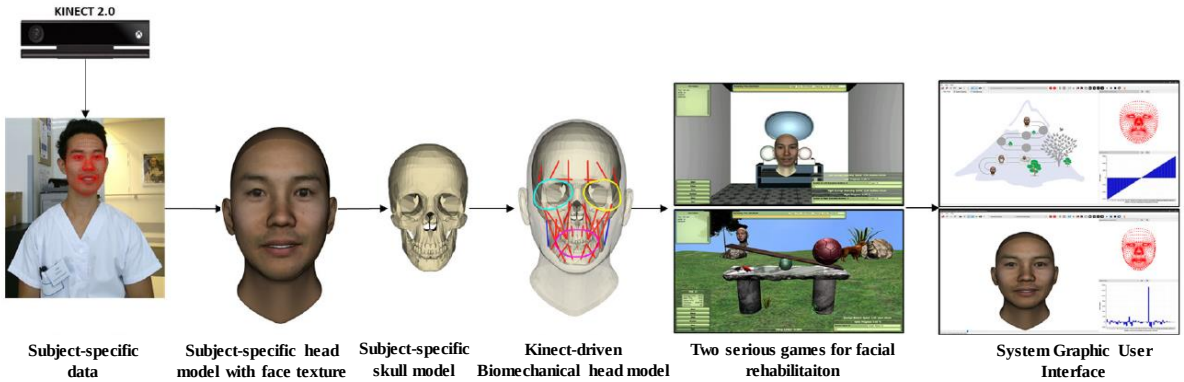


Figure 12. A clinical decision support system proposed by Nguyen 2020 for facial palsy rehabilitation [105].

Firstly, a subject/patient-specific 3D head model, as well as face texture, was generated from the 3D data collected by using Kinect sensor version 2.0. Secondly, a skull model was regressed along with the facial muscle network. Two serious games for facial rehabilitation

oriented with smiling and kissing practices were developed using the Kinect-driven biomechanical head model. Finally, a graphic user interface was developed for the users.

To sum up, facial palsy patients or patients under facial transplantation have facial dysfunctionalities and abnormal facial movements. This negatively affects the personal and professional life of involved patients. The recovery of the symmetric face with balanced functionalities requires a complex rehabilitation process in which patients must practice patient-specific facial movements. Most traditional facial rehabilitation has mainly been based on a mirror approach to monitor the visual qualitative feedback from the rehabilitation activity. In fact, patients use their distorted features in the mirror as a reference to teach themselves the correct expressions, making it difficult to practice appropriately. This strategy is ineffective and subjective without any additional feedback. On the other hand, machine learning and deep learning have been extraordinarily applied in many fields including healthcare settings. However, according to our knowledge, none of these studies has been applied to target the facial rehabilitation process. Moreover, despite the fact that a variety of computer-aided decision support systems have been produced for treatment and rehabilitation, very few systems have been established for facial mimic rehabilitation. The system provided by Nguyen 2020 was the only system that provides quantitative and objective bio-information for facial rehabilitation. But the system has not used biomechanical knowledge of the physical-based model. Furthermore, the use of the system in clinical routine practice still remains challenging due to **the lack of building 3D information** from images or depending strongly on the selected cameras. The system is also limited due to **the lack of analyzing the face in terms of expression recognition and symmetry.** These are of importance to better quantify the deformation level of patients and optimize the facial rehabilitation program. Moreover, the current rehabilitation system is also limited by a lack of patient-specific knowledge about **muscles driving facial motions with emerging biomechanical knowledge**. This is an important indicator for guiding patients to practice rehabilitation exercises.

To overcome the limitations, the objective of my Ph.D. thesis was to build a clinical decision support system for improving the facial rehabilitation process that uses portable devices coupled with biomechanical knowledge and artificial intelligence technologies.

1.2. Objectives

The objectives of the Ph.D. project are to propose **a framework** (REHAB_DEEPFACE) and **innovative engineering solutions** toward a next-generation **computer-aided decision support system for facial analysis and rehabilitation**. The target clinical application of this project is facial palsy disorder.

The thesis provided four main contributions to solve some solutions of REHAB_DEEPFACE:

1) The fast reconstruction of the 3D face shape from a single image using deep learning approaches (chapter 3).

2) The improvement of facial expression recognition using 3D point sets and geometric deep learning (chapter 4).

3) The face symmetry analysis based on novel descriptors (chapter 5).

4) The proposition of a novel modeling framework for learning facial motion by coupling reinforcement learning and finite element modeling for facial motion learning and prediction (chapter 6).

The framework of REHAB_DEEPFACE is illustrated in Figure 13. Firstly, 3D face of a patient will be reconstructed from a single 2D image. Secondly, recognition of facial behaviors will be done using deep learning models (e.g. geometric deep learning). Finally, a reinforcement learning model will be developed and coupled with a finite element model of the face for facial motion learning. This will allow exploring skin deformation and muscle contraction behaviors and their effect to optimize the rehabilitation movement.

In perspective, clinical evaluations of the proposed solutions have to be performed after the progress of the project.

1.3. Thesis organization

Besides this chapter, the thesis is organized as the followings:

- **Chapter 2:** *State-of-the-Art: 3D face reconstruction, 3D facial expression recognition with deep learning, facial symmetry analysis, facial movements for finite element model of the face.* The existing approaches and limitations of 3D face reconstruction from a single picture, facial analysis such as 3D face expression recognition and facial symmetry analysis, and facial motion learning by coupling reinforcement learning and the finite element model of the face are presented.

- **Chapter 3:** *3D face reconstruction from a single image using different deep learning approaches for facial palsy patients.* The primary goal of reconstructing a patient's 3D face is to provide trustworthy feedback for clinical decision support. A methodology is proposed to reconstruct the facial palsy patient's 3D face shape models in natural and mimic postures from a single 2D image.
- **Chapter 4:** *Enhanced facial expression recognition using 3D point sets and geometric deep learning.* In human communication and human-computer interaction, facial expressions are crucial. A new class of deep learning called geometric deep learning was used to recognize facial expressions directly on 3D point cloud data. The accuracy of the recognition based on the hyperparameter tuning process and cross-validation methods will be presented as well.
- **Chapter 5:** *Facial symmetry analysis based on novel shape descriptors between two different populations.* The balance and equality of face anatomy are referred to as facial symmetry. A new descriptor for face shape will be retrieved, and the symmetry of the face will be analyzed.
- **Chapter 6:** *Reinforcement learning coupled with finite element modeling for facial motion learning.* Understanding facial motion mechanism remains a scientific and clinical challenge to help the involved patients to recover symmetrical movements and normal facial expressions. I will provide a new framework for learning facial motion during facial expression motions that combines reinforcement learning with a finite element model of the face. This chapter will go through the process of exchanging data, training, and predicting face movements.
- **Chapter 7:** *General discussion.* A summary of the work and main contributions will be discussed.
- **Chapter 8:** *Conclusions and perspectives.* The final chapter gives conclusions and perspectives.

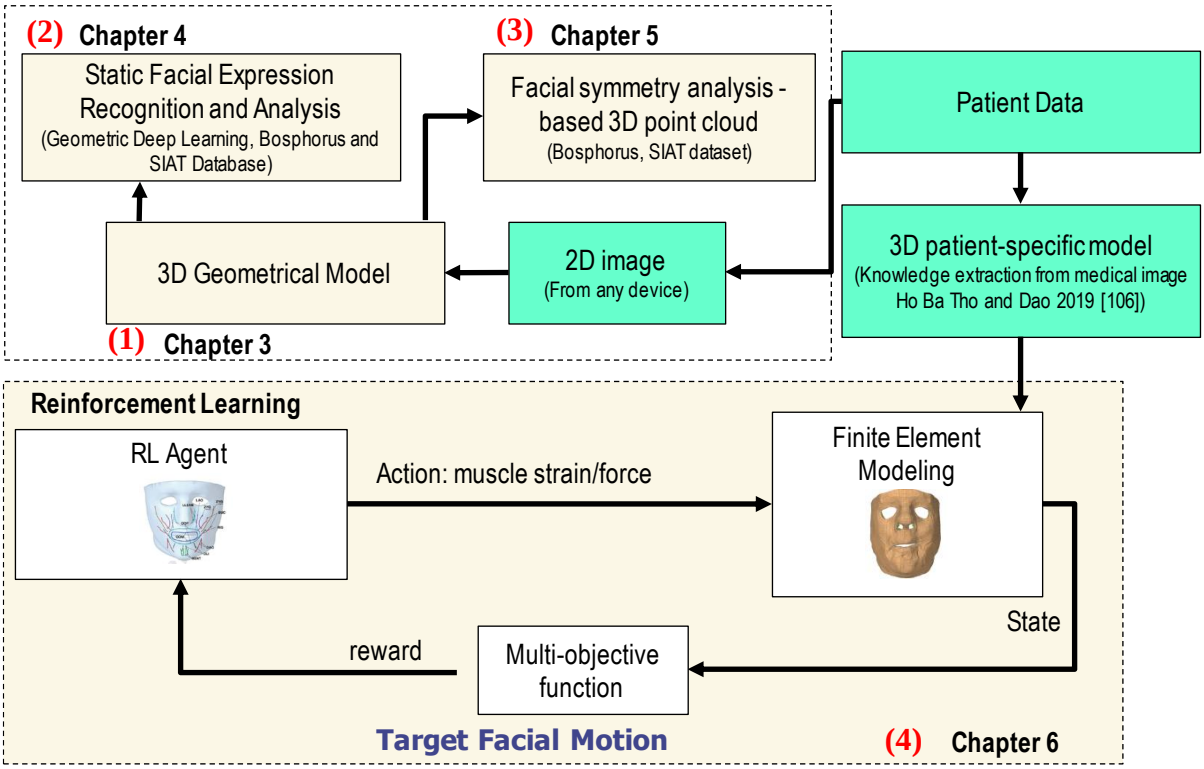


Figure 13. A framework (REHAB_DEEPFACE) for the project from 2D patient data → (1) reconstructing a 3D geometric model of the face (chapter 3) → (2) analyzing the face in terms of facial expression (chapter 4) and (3) facial symmetry (chapter 5) → (4) facial learning motion by coupling between reinforcement learning and finite element model of the face (chapter 6).

Chapter 2:

State-of-the-Art: 3D face reconstruction, 3D facial expression recognition with deep learning, facial symmetry analysis, facial movements for finite element model of the face

An accessible decision support system should help patients in both the diagnosis and rehabilitation process. A **3D face of the patient** is needed so that more analysis on the face could be done automatically. The analysis often includes quantifying **facial expressions** and **facial asymmetry and symmetry**. So that the level of the severity of facial palsy patients can be identified. Then the rehabilitation process can be planned. Besides, **understanding facial motion mechanisms** by knowing which muscle is responsible for what movements could guide patients to practice rehabilitation exercises properly.

In this chapter, I introduce as well as analyze achievements of the state-of-the-art for several domains related to the thesis project such as 3D face reconstruction, 3D facial expression recognition, facial symmetry analysis, and facial movement learning.

2.1. 3D face reconstruction from a single image

2.1.1. Introduction

The patients, who are involved in facial palsy or facial transplantation, suffer facial dysfunctionalities and abnormal facial motion. This is due to the altered facial nerve and facial muscle systems [31], [107]. This leads to unwanted facial movements such as dysfunctionalities of speaking, eating, and unnatural relaxation of mouth corner drop, eyelids closures, and asymmetrical facial expressions [21], [108]. Recently, computer-aided decision systems have been developed to provide objective and quantitative indicators to better diagnose and optimize rehabilitation programs [109]. The 3D reconstruction of an accurate face model is essential to provide reliable feedback. The 3D face is currently achieved by using medical imaging [110] and different sensors like Kinect or 3D scanners [111]–[113]. Thus, this allows analyzing the face with external (i.e. face deformation) and internal (i.e. facial muscle mechanics) feedback for the diagnosis and rehabilitation process of facial palsy and facial transplantation patients [33].

2.1.2. Collecting 3D facial database

In the past few decades, facial analysis has attracted great attention due to its numerous exploitations in human-computer interaction [114], [115], animation for entertainment [116]–[118], and healthcare systems [15], [119], [120]. Facial analysis from 2D images remains a challenge due to variation in poses, expressions, and illumination. 3D information can be used in order to cope with these variation problems.

3D facial data can be acquired from medical imaging [121], [122]**, 3D scanners** [112], [113]**, stereo-vision systems** [95] **, or RGB-D devices as Kinect**. Specifically, facial reconstruction produces that are pretty near to the original skin were conducted based on extracting values of the soft tissue thickness at distinctive landmarks and the original skull from *magnetic resonance images* (MRI) [121]. Different 3D shapes of the face can be generated from the same skull model by adjusting the value of the soft tissue thickness regarding the positions of selected facial landmarks. Two stages of the facial reconstruction process are conducted using this method. In the first stage, initialization of the deformable model was inferred using the reference skin and the reference skull. The previously initial skin model was refined in the second stage using the means of a 3D deformable model built from simplexes meshes based on selected facial landmarks. The 3D shape of the face with various facial expressions can also be successfully captured with high accuracy using *3D scanners* such as structured-light Inspeck Mega Capturor II 3D [113]. This system requires subjects sitting at 1.5 meters away. The resolution is $0.3 \times 0.3 \times 0.4$ mm in corresponding x, y, and z directions. Or it can be captured by a *3D face imaging system* [112]. This system projects a random light pattern onto the person and records his or her shape using multiple digital cameras mounted at various angles and precisely synchronized. By using these cameras, both the shape geometry and surface texture of the face of the person were precisely captured. Another way used a contactless *Kinect sensor* [111]. Using superior data from

Kinect, the system can reconstruct subject-specific geometrical heads and face models with texture information, as well as animate rigid and non-rigid facial mimic motions in real-time.

The use of medical imaging leads to a very accurate 3D model but this is not appropriate for an easy-to-use, cheap and portable system. **The use of depth cameras like Kinect** can lead to a reasonable accuracy level while keeping the cheap cost, easy-to-use, and portable requirements. But the **developed system depends strongly on the selected sensors**. In fact, this could alter the future applicability due to stopped production like in the case of the Kinect V2 camera. Thus, it is necessary to have a more flexible and open method to build 3D information rather than using specific scanning devices.

2.1.3. Application of 3D face reconstruction

Numerous applications have been established by 3D shape reconstruction from 2D images. In the computer animation field, it helps to create 3D avatars from images. This can also be used in entertainment fields (e.g. virtual reality or gaming applications) when it is necessary to embed the user avatar into the system [124]. Chien-Hsu Chen et al. [124] (2015) proposed to use of an augmented reality-based self-facial modeling system. The system overlays 3D animation of participant faces for six basic facial expressions, allowing them to practice emotional assessments and social skills. The virtual avatars' 3D head and face models were created to suit patients. And then the system was applied for patients to practice emotional and social skills by allowing the virtual avatar models to perform six fundamental facial expressions. Additionally, a reconstructed 3D face provides biometric features for security purposes such as human identification [125], [126], and human expression recognition [127]. In fact, the face of a person can be used as particular biometric evidence along with other biometric information such as what a person has (e.g. iris, fingerprint, retina, etc.) or produces (e.g. gait, handwriting, voice, etc.) [125]. Biometric facial recognition is an appealing biometric technique because it relies on the same identifier that people use to differentiate one person from another: **the face**.

2.1.4. 3D face reconstruction methodology

Various methods have been developed to estimate 3D face shapes from one image or multi- images [128]. Three distinguish approaches have been applied for reconstructing 3D shapes from 2D information namely 1) statistical model fitting method, 2) photometric method, and 3) deep learning method.

2.1.4.1. Statistical model fitting approaches

The first approach uses a prior statistical 3D facial model to fit the input images [129]–[131]. In fact, 3D face reconstruction from 2D images is an ill-pose problem. It needs some types of previous knowledge. In order to find the solution, statistical 3D face models are preferred methods to incorporate this previous knowledge since they encode facial geometric variations. The 3D morphable model is a statistical 3D face model built from a set of 3D scans of heads. This model includes both the shape and the texture of the face. Firstly, they collected a large number of faces by a 3D scanner. Each face was presented as a set of cloud points in the 3-dimensional space $s = [x_1, y_1, z_1, \dots, x_n, y_n, z_n]^T$. The structure and order of

these points are the same for every face scan. Secondly, all these facial shapes were decomposed into a linear combination of the mean shape ($\bar{\boldsymbol{s}}$) and the shape bases ($\boldsymbol{s}_i$).

$$\boldsymbol{s} = \bar{\boldsymbol{s}} + \sum_{i=1}^{m} \boldsymbol{\alpha}_i \boldsymbol{s}_i$$

Similarly, all the facial textures were also decomposed into a linear combination of the mean texture ($\bar{\boldsymbol{t}}$) and the texture bases ($\boldsymbol{t}_i$)

$$\boldsymbol{t} = \bar{\boldsymbol{t}} + \sum_{i=1}^{m} \boldsymbol{\beta}_i \boldsymbol{t}_i$$

The mean shape and mean texture were evaluated from all 3D faces from 3D scanners. Finally, the 3D face reconstructed from one image will be performed by estimating the shape (α) and the texture (β) coefficients.

There are several existing 3D statistical models in the last few decades. For example, Blanz and Vetter (1999) created a 3DMM in UV space from 200 young adults including 100 females and 100 males [132]. The well-established Basel Face Model (BFM) [133] was built from 200 subjects (100 males and 100 females with an average age of 24.97 years old from 8 to 62) with most of the subjects being Caucasian. The Surrey Face Model (SFM) [134] was constructed from 169 subjects. It includes a diversity of both ages (from 0 to more than 60 years old) and ethnicity (60% of the subjects are Caucasian, 20% of the subjects are Eastern Asian, 6% of the subjects are Black Asian, and 14% of the subjects are other ethnicities such as Arabic, South Asian, and Latin). Two other 3DMM models are the Large Scale Facial Model (LSFM) [135] and the Liverpool-York Head Model (LYHM) [136], [137]. Both models were built from a wide range of ages and balanced gender (with ratios of males/females are 48/52 for LSFM and 50/50 for LYHM). All of these models were built in the neutral position, while other methods built 3DMM which capture geometric variations related to expressions. Several examples are FaceWarehouse model [138], CoMA model [139], FLAME model [140], BFM 2017 model [141], and Multilinear wavelet model [142]. In particular, the FaceWarehouse model was built by Cao et al. [138] (2014) from depth images of 150 participants. Each has 20 different expressions and the age range is between 7 and 80 years old. More existing 3DMMs are shown in Table 1.

Table 1. Existing 3DMM

3DMM	# subjects	% males /females	Age range	Ethnicity	Expression	# cites
Blanz Vetter 1999 [132]	200	50/50	Young adults	-	No	5272
BFM 2009 [133]	200	50/50	8-62	Most Caucasian	No	1010
FaceWarehouse [138]	150	-	7-80	Various	Yes	750
SFM [134]	169	-	Wide range	60% Caucasian	No	202
CoMA [139]	12	-	-	-	Yes	154
FLAME [140]	3800	48/52	Wide range	Wide range	Yes	198
LSFM	9663	48/52	Wide range	82% white	No	187
BFM 2017 [141]	200	50/50	-	-	Yes	18
LYHM [136], [137]	1212	50/50	Wide range	-	No	89
Multilinear Wavelet model [142]	99	-	-	-	Yes	23

Afterward, by identifying parameters of the linear combination of 3D statistical model bases that best matches the provided 2D image, a new 3D face can be reconstructed from one or more images.

2.1.4.2. Photometric approaches

The second approach is based on the photometric stereo. The method is suitable for multiple images. The method combines a 3D template face model with photometric stereo algorithms to compute the surface normal of the face [143], [144]. The surface normal and the lighting parameters from a set of images are usually estimated by an assumption of a Lambertian reflectance model for each pixel (with coordinate (x, y)) of the input image $\mathbf{I}$ as

$$\mathbf{I}(x, y) = \mathcal{A}(x, y)\mathbf{l}^T\mathbf{n}(x, y)$$

where at the pixel location (x, y), $\mathcal{A}$ represents the albedo, $\mathbf{n}$ represents surface normal with light source vector $\mathbf{l}$. However, due to the unconstrained nature of images, the lighting source is normally unknown so the method is purely based on the albedo. Thus, additional prior knowledge should be added. This prior knowledge can typically be a set of different images under different poses and angles, lighting conditions, and expressions from the same subject. Or it uses a single image with a combination of 3DMMs or a 3D template model.

Various approaches have been proposed for reconstructing the 3D face of a person using photometric methods with multiple images. Kemelmacher-Shlizerman and Seitz [139] (2011) used multiple images to reconstruct the 3D face. They applied the matrix decomposition for $\mathbf{M} \in \mathbb{R}^{n\times d}$ with each column corresponding with a frontal image of the same person. The

method outputs a matrix $\mathbf{\Gamma} \in \mathbb{R}^{n \times d}$ including the lighting parameters and a matrix $\mathbf{S} \in \mathbb{R}^{4 \times d}$ including the albedo and the surface normal. Another example of Suwajanakorn et al. [140] (2014) reconstructed the 3D shape of the same person from each frame from a video sequence using a 3D optical flow technique paired with shading. On the other hand, Snape et al. [141] (2015) proposed to decompose a matrix $\mathbf{M}$ with each vectorized image that possibly contains more than one person at a column into $(\mathbf{\Gamma} * \mathbf{C})\mathbf{S}$. Where $*$ represents the Khatri-Rao product and $\mathbf{C}$ matrix including the shape parameters associated with the identity of subjects inside the input images. This method is able to reconstruct multiple subjects at once. **All these methods require a large set of images**. In contrast, Zeng et al. [142] (2017) only used 3 images (including one frontal and two side views) that cooperated with prior knowledge as a collection of reference meshes. At the beginning step, 3 face models were coarsely reconstructed for three input images using a single-image reconstruction approach proposed by [143]. The first model deals with *shading term* that penalized the variation between the initial and the new estimations. The second model deals with *consistency term*, which is associated with multi-view to be in agreement for all the reference shapes. The third model is *smoothness term* to make sure smooth transitions in depth.

On the other hand, other studies reconstructed the 3D shape of the face of a person using only one single image combined with prior knowledge. Three alternative approaches were proposed. 1) The first approach fits a pre-designed template face model to the input image [143]. 2) The second approach trains a face model combining shape and illumination parameters [145]–[147]. 3) The final approach fits a 3DMM to produce a coarse model that is then refined [148]–[150]. For the first approach, Kemelmacher-Shlizerman and Basri [143] (2011) reconstructed the 3D face of a person based on a template model. This method estimated each of three elements including the surface normal, albedo, and depth map alternatively by fixing the two remaining. In particular, the spherical harmonic parameters γ were estimated by fixing the albedo and the normal of the template model, while fitting the reference shape into the input image. The depth map from the input image is computed by using pre-computed γ and albedo parameters. Finally, the albedo was recovered using pre-computed γ and the depth map. In the training face model methods, Lee and Choi [145] (2011) modelized the input image ($\mathbf{I} \in \mathbb{R}^{n_x \times n_y}$) by using spherical harmonics to approximate a Lambertian reflectance model, resulting in

$$\mathbf{I} = \mathbf{F} \times_3 \boldsymbol{l}^T$$

where $\mathbf{F} \in \mathbb{R}^{n_x \times n_y \times n_l}$ describes parameters related to albedo and normal of the surface, $\boldsymbol{l} \in \mathbb{R}^{n_l}$ represents the light source. The surface property matrix $\mathbf{F}$ was then parameterized as a function of personal identity $\tau \in \mathbb{R}^{n_{id}}$ using the mean $\bar{\mathbf{F}} \in \mathbb{R}^{n_x \times n_y \times n_l}$ and the bases function $\mathbf{T} \in \mathbb{R}^{n_x \times n_y \times n_l \times n_{id}}$, which results in $\mathbf{F} = \bar{\mathbf{F}} + \mathbf{T} \times_4 \tau$. Finally, $\mathbf{I}_{new}$ can be reconstructed by estimating $\boldsymbol{l}$ and τ which minimize the reconstruction error between $(\bar{\mathbf{F}} + \mathbf{T} \times_4 \tau) \times_3 \boldsymbol{l}^T$ and $\mathbf{I}_{new}$.

The third approach uses one single image paired with a 3DMM to guarantee that the reconstruction is generally realistic [148]–[150]. In general, they solely employed photometry-based approaches to improve a rough 3D face approximated by fitting a 3DMM.

Especially, without the assumption of uniform surface albedos, a robust optimization approach was developed to accurately calibrate per-pixel illumination and lighting direction [148]. The input images are then semantically segmented using a customized filer along with the geometry proxy to adjust hairy and bare skin areas.

2.1.4.3. Deep learning approaches

The third approach uses deep learning to learn the shape and appearance of the face by training 2D-3D mapping functions [151], [152]. The method encodes prior knowledge of the 3DMM into the weights of the deep neural network. Directly learning the 2D-3D mapping function is challenging due to lacking ground truth 3D face model. Thus 3 different approaches including the **training dataset**, the **learning framework**, and the **training criterion** have been proposed to cope with this issue regarding the learning process association.

Regarding **the training dataset**, having a huge number of 3D face models paired with the 2D images is most likely impractical. Numerous researches focused on enhancing the database by constructing the synthesis data for the training data set from existing 3DMMs. These methods include 3 different strategies:

1) *The fit and render method*, which fits an existing 3DMM into real facial images and then uses these 3D reconstructed face models to render synthetic images [153]–[155].

2) *The generate and render method*, which randomly generates 3D face models by adjusting the 3DMM parameters and then renders synthetic images under different conditions relating to poses, lighting, expressions, etc. [156], [157].

3) *Self-supervision training method*, which can avoid the requirement of pairs of 2D-3D ground truth data as well as the synthetic images. The method modifies the training network with an additionally rendering layer at the end of it. Then the training end-to-end process was spawned by minimizing the loss function generated from the input and rendered images *[158]–[161]*.

According to the **learning framework method**, various different learning strategies have been proposed for reconstructing the 3D face of a person. Those include:

1) *The use of a single neural network for a single pass training*: Convolutional neural network is a regular choice. The main idea of this method is to regress the parameters including shape, expression, and texture parameters of the existing 3DMM from a single input image [162]–[164].

2) *The training of neural networks in an iterative way*: This can be conducted either by focusing on iteratively enhancing the synthetic training set or based on iteratively refining the previous iteration's outcome. Kim et al. [165] (2022), for example, rendered synthetic images using parameters predicted based on the trained neural network from a real image. Then these synthetic images were added to the training set in each iteration. As the result, after each

iteration, the training dataset was augmented by combining the data generated by the training network.

3) *The use of encoder-decoder architecture*: The method divides the network into two separate parts. The encoder part makes the dimensional reduction of the input image to find new representative features. While the decoder part makes use of the new representative features to reconstruct the 3D facial geometry of a person [166]–[169].

4) *The use of generative adversarial networks*: The method trains two distinct networks namely generator and discriminator to learn the distribution of 3D faces from ground truth data, which aims to obtain more realistic 3D faces. For example, Tu et al. [170] (2020) trained a generative adversarial network to regress the parameters of a 3DMM. Gao et al. [171] (2020) reconstructed a realistic 3D facial geometry of a person from a decoder network trained by the discriminator network.

5) *The use of multiple networks*: The method uses multiple networks with each network being responsible for a specific sub-task to determine different parameters. For example, Tewari et al. [172] (2019) and Bhagavatula et al. [173] (2017) extracted features from input images by a CNN and then fed these features into multiple networks for predicting different parameters involving shape and albedo. On the other hand, Fan et al. [174] (2021) fit 3DMM into the input facial image to reconstruct a personalized template model, which was then refined using the 50-layer ResNet [175]. Wang et al. [176] (2020) also regressed the 3DMM parameters to construct a coarse 3D model, which was then refined by another refine network. Dou et al. [177] (2017) trained multiple networks in parallel. This is able to estimate different parameters by adding different branches of sub-CNN-network to the main network of VGG-Face Network [178]. This sub-CNN-network predicts parameters related to expression, while the main network predicts parameters related to identity.

The training criterion approach is regularly related to the loss function, which represents the variation between the output and the provided ground truth, and should be minimized during training the network.

To sum up, the 3D reconstruction of an accurate face model is essential to provide reliable feedback. Thus, this allows analyzing the face with external (i.e. face deformation) and internal (i.e. facial muscle mechanics) feedback for the diagnosis and rehabilitation process of facial palsy and facial transplantation patients. Previous studies have mainly been based on three different approaches. The first approach that **fits a 3DMM** to 2D images is mainly based on estimating the parameters of the linear combination of the model bases. The fitting process is driven by corresponding landmarks or texture edges corresponding to 2D-3D points. The second approach reconstructs 3D face shape from one or multiple images using **photometry** was mainly based on a Lambertian illumination model that transforms an input image into albedo, normal, and lighting to recover the surface normal. The final approach is based on **deep learning** to learn the shape and appearance of the face by training 2D-3D mapping functions.

These approaches can lead to very good accuracy levels for 3D face reconstruction and 3D subject-specific face reconstruction. However, the 3D face reconstruction of facial palsy patients is still a challenge and this has not been investigated. The objective of Chapter 3 will propose a methodology to reconstruct the 3D face shape models of the facial palsy patients in natural and mimic postures from one single 2D image. Then, based on the outcomes, the best method will be selected and implemented into our computer-aided decision support system for facial disorders.

2.2. Facial expression recognition

2.2.1. Introduction

Facial expression recognition (FER) plays an important role in numerous applications. Recognizing human expression is of helps humans in face-to-face communication in the interpretation of the other's intentions. Mehrabian (1975) illustrated that the part related to facial expressions of a speaker can account for 55% of the interpretation in the conversations, while the verbal part (i.e. part relates to words) and the vocal part (i.e. part relates to the sound) contribute only to 7% and 38%, respectively [8]. There are also a wide range of applications of facial expression recognition [179] in the human-computer interaction [115], especially for the virtual reality and augmented reality system [124], [180], and healthcare systems (e.g. facial nerve grading [181], [182]). More specifically, facial expression recognition can be used for *developing software*, where it was a part of measuring user satisfaction while using the software [183]. In the *education area*, facial expression recognition assists to monitor feedback from both lecturers and learners [184], [185]. This can also evaluate the dynamism of class and effectiveness of teaching in a distance teaching environment such as tele-teaching or virtual learning. In the *medicine and healthcare area*, facial expression helps to monitor the emotion and expression perception of involved patients who suffer neuro-psychiatric disorders. Automatic facial expression recognition offers an objective and quantitative process of monitoring patients' treatment feedback, and rehabilitation therapy [186]–[188]. Facial expression recognition can also be useful in *security applications* such as security surveillance systems or biometric systems related to facial recognition. Integrating facial expressions could make these systems able to detect malicious intention [189], [190]. *In marketing*, facial expression recognition can be used for capturing facial behaviors of users while trying a product's sample [191] as well as keeping track of client interest and advertisement approval [192], [193].

There are several important terminologies for FER research that are given below:

The *facial action coding system* (FACS), proposed by Ekman and Friesen [194] (1978), tracks changes in facial muscle contractions during a person's expression. The contractions of individual facial muscles, known as *action units* (AUs), are encoded by FACS and represent discrete instantaneous changes in face appearance [195].

Facial landmarks (FLs) are key points in the face such as points on eyebrows, eyes, nose, and mouth (Figure 14).

Basic expressions (BEs) comprise 6 basic expressions (e.g. anger, disgust, fear, happiness, sadness, and surprise) and one neutral expression (Figure 15).

Compound expressions (CEs) are expressions that combine two basic expressions. Six basic expressions and a neutral expression, 12 compound expressions, and three others

including appall, awe, and hate are the 22 most typical expressions performed by humans [196].

Micro-expressions (MEs) are involuntary facial movements that are more spontaneous and delicate in a short time.

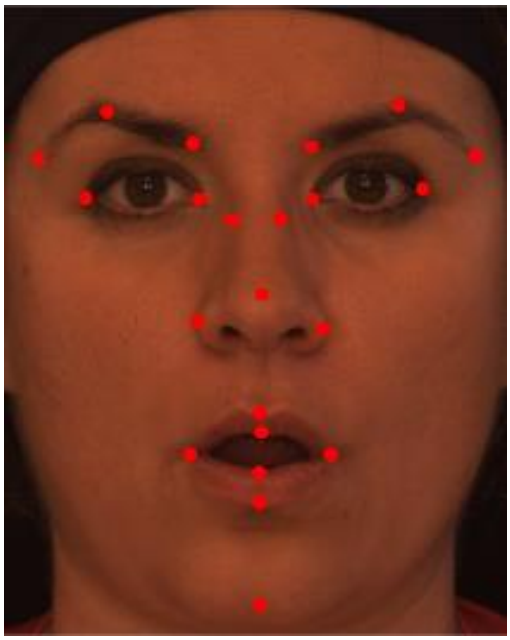


Figure 14. Facial landmarks include salient points in the interesting regions in the face [113].

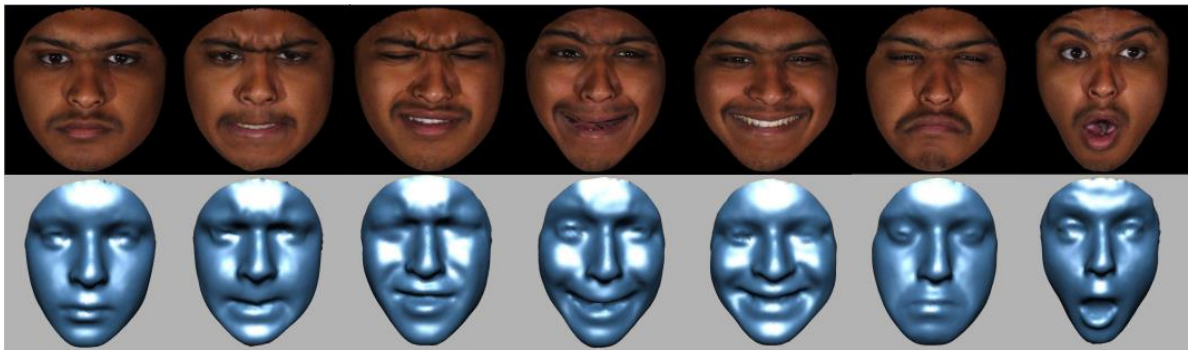


Figure 15. Six basic facial expressions and a neutral position from BU-3DFE database [112]

2.2.2. Existing facial database

Various sources of databases related to facial expression have been recently published including 2D face and 3D face scans. These databases were captured in both controlled (in the laboratory) and uncontrolled (in real-world) conditions. Two databases namely Cohn Kanade (CK) [197] (2000) and Cohn Kanade extension (CK+) [198] (2010). CK includes 97 subjects (the age ranges from 18 to 30 years old, female subjects count 65%, and male

subjects count 35%) with a total of 486 sequences of 2D images. CK+ extended the existing CK database to increase to 123 subjects (the age ranges from 18 to 50 years old, 69% of subjects are female, and the remaining 31% of subjects are male). Each subject performed 6 basis expressions (6E) and one neutral position following the FACS code. A 3D face database namely Bosphorus [113] (2008) comprises 105 subjects (60 males and 45 females, 18 subjects wore mustaches/beards). Each subject was scanned 31 to 54 times under different expressions, and head poses. Binghamton university 3D facial expression (BU-3DFE) [112] (2006) contains 100 subjects (56 females and 44 males, the age ranges from 18 to 70 years old, various ethnicities such as White, Black, Middle-east Asian, East-Asian, Indian). Each subject was captured in different expressions and poses. The metadata comprises 3D face scans as well as associated facial textures with a total of 2500 facial scan models. Two other databases were developed from Binghamton university namely Binghamton university 3D dynamic facial expression (BU-4DFE) [199] (2008) and Binghamton-Pittsburgh 3D dynamic spontaneous facial expression database (BP4D) [200] (2014). The former includes 606 sequences of 3D facial scans (100 frames for each sequence) of 101 subjects (58 females and 43 males with various ethnicities) captured with six basic expressions. The latter includes 41 subjects (23 females and 18 males, the age ranges from 18 to 29 years old, with various ethnicities) with each subject conducting spontaneous 3D facial expressions. Another worth mentioning database namely the SIAT-3DFE database [201] (2020) was also applied for facial expression recognition. This high-resolution database was acquired by two structured light scanner systems called CASZM-MVS600. The data consists of 8000 3D facial scan models from 500 Asian volunteers aged from 18 to 50. Each participant was scanned 16 times with 4 basic facial expressions such as neutral, happiness, sadness, and surprise. More benchmark facial databases were shown in Table 2.

Table 2. Several existing benchmark facial databases. E = expression, N = Neutral

Database	# subjects	Total face	Type	Environment	Expression
CK [197]	97	486	2D	Laboratory	6E + 1N
CK+ [198]	123	593	2D	Laboratory	6E + 1N
Bosphorus [113]	105	4888	3D	Laboratory	6E + 1N
BU-3DFE [112]	100	2500	3D	Laboratory	6E + 1N
BU-4DFE [199]	101	606	4D	Laboratory	6E + 1N
BP4D [200]	41	-	4D	Laboratory	6E + 1N
SIAT-3DFE [201]	500		3D	Laboratory	3E + 1N
FER2013 [202]	-	35,887	2D	Internet	6E + 1N
MMI [203]	25	740I, 2900V	2D	Laboratory	6E + 1N
AFEW 7.0 [204]	-	1809V	2D	Movie	6E + 1N
4DFAB [205]	180	1.8 million	4D	Laboratory	6E + 1N
EmotioNet [206]	-	450,000	2D	Internet	23E or compound expressions
JAFFE [207]	123	213	2D	Laboratory	6E + 1N
Oulu-CASIA [208]	80	2880	3D	Laboratory	6E

2.2.3. Facial expression recognition framework

A facial expression recognition or classification contains three main tasks including preprocessing data, feature extraction, and classification or recognition (as presented in Figure 16).

The first crucial part of FER is **pre-processing data**. This step has a substantial impact on the performance of both traditional machine learning as well as deep learning algorithms. Depending on the use of the input data (2D images, 3D face scan, sequences of images), several techniques could be applied for pre-processing data process. 1) The first technique is face detection helps to identify the facial region in the image applied for the method using 2D images as input [209]–[213]. 2) Facial landmark detection determines key points in the face such as the eyes, eyebrows, nose, mouth, and lip [214]–[216]. 3) After that, facial normalization was applied to minimize the feature dependence related to space, pose as rotation, lighting conditions as brightness and illumination, occlusion, etc. [217], [218]. 4) And finally, data augmentation could significantly improve machine learning and deep learning performances. Because it enlarges the data size through several techniques such as rotating, flipping, scaling, cropping, resampling, etc. [219]–[222].

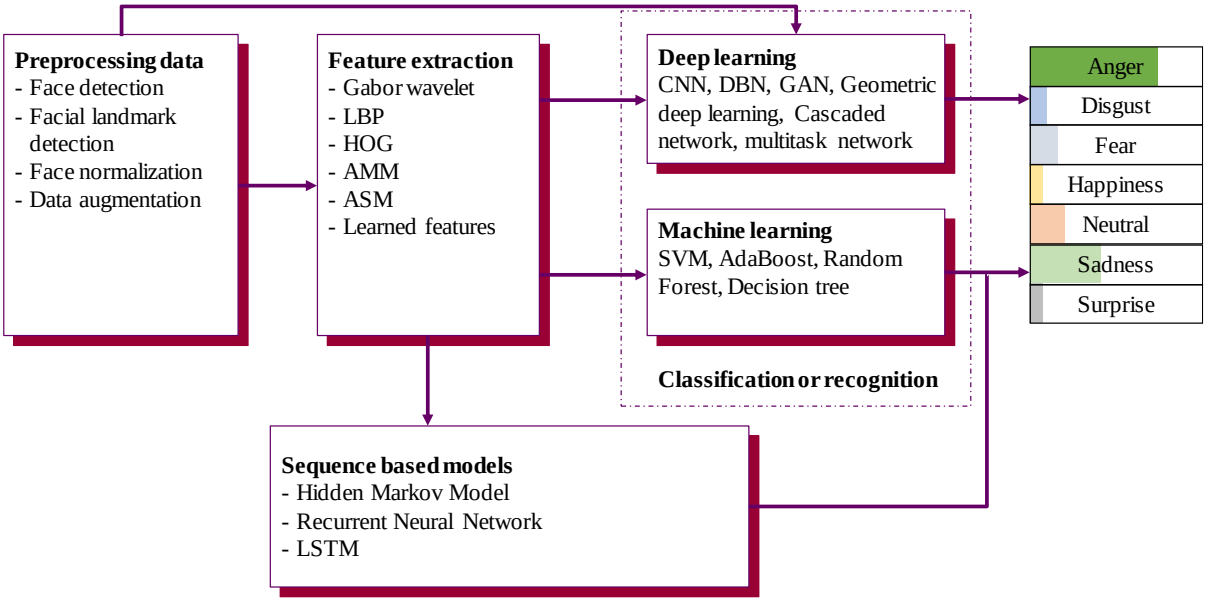


Figure 16. Facial expression recognition usually contains three main tasks: preprocessing data, feature extraction, and classification or recognition.

The second main part of FER is **feature extraction**, which aims to extract useful information from input data. This information can maximum interclass variation, while minimum intraclass variation to discriminate different expressions. 1) Hand-crafted feature extraction, 2) learning feature extraction, and 3) hybrid feature extraction are popular techniques to extract features for a FER problem. *Hand-crafted feature extraction* includes appearance-based features and geometric-based features. Appearance-based features capture changes in the facial image related to the shape, texture, color, or key points in the face. It

then presents these changes using one single representative vector. Geometric-based features, on the other hand, capture the displacements during different expressions of pre-defined facial landmarks. Regarding the appearance-based feature extraction, numerous methods have been used such as Gabor Wavelet [223], local binary pattern (LBP) [224], histogram of oriented gradient (HOG) [225], principal component analysis (PCA) [226], scale-invariant Fourier transform (SIFT) [227], [228]. These appearance-based features, in general, are able to catch transient alterations in the face such as wrinkles, furrows, bulges, etc. However, these types of features are vulnerable to changes in lighting conditions and image quality. Geometric-based features use an active appearance model (AAM) [229] to match a collection of model points to an image. Or it uses an active shape model (ASM) [230] to match the shape and texture of objects to an image, to track a set of pre-defined face landmarks. In spite of being not influenced by lighting conditions, simple to track as well as working well for several expressions, they are insufficient for expressions that do not result in the pre-defined facial landmark displacement. The second type of feature namely *learned-based features* extracted from neural networks and deep learning networks. Other networks can also be used like convolutional neural networks, recurrent neural networks, and geometric deep learning networks. However, due to the lack of a large database for training networks, the learned features might not sufficiently capture discriminative information for FER. Another type of feature for FER fuses different features to get prediction namely *hybrid-based features*. Due to combining other features' advantages, this type of feature could dramatically improve the recognition rate.

Many traditional machine learning algorithms have been used for **recognizing or classifying** facial expressions such as support vector machine [231], [232], adaptive boosting [233], random forest [234], [235]. In general, traditional machine learning algorithms require hand-crafted feature processing for FER. Numerous classes of deep learning algorithms were also applied for FER such as CNN class with LeNet [236], AlexNet [237], ResNet-50 [238], [239], GoogLeNet [240], [241], VGGNet [242], [243], recurrent neural network (RNN) [244]–[246], Dynamic Bayesian Network (DBN) [247], GAN [248]–[251], multitask learning systems [252], [253], geometric deep learning [127]. Deep learning algorithms for FER might avoid hand-crafted feature extraction, but they demand a large scale of databases for training. Besides, the tuning process to find the optimal hyperparameter for the training is also high computational cost and time-consuming.

2.2.4. 3D facial expression recognition based on 2D facial images

Most existing facial expression recognition in the past thirty years has primarily been based on 2D image processing. This can be divided into methods using conventional machine learning or deep learning [7] (Figure 17). Various **conventional machine learning** approaches extract facial appearance features, facial geometric features, or/and a combination of both (Figure 18). For example, Happy et al. [254] (2012) extracted appearance features in the global face region using a local binary pattern (LBP) histogram, then applied PCA for classifying facial expressions. Due to the lack of local variations in the face, the accuracy tends to deteriorate. Ghimire et al. [255] (2017) divided the face into different sub-local

regions and extracted appearance features for these sub-local regions, then applied a support vector machine to recognize facial expressions. In other research, Ghimire and Lee [256] (2013) recognized expression from sequences of images by extracting two different geometric features related to the angles and distances created from 52 pre-designed facial landmarks. Euclidean distances and angles were first computed from each pair of facial landmarks of a frame and then subtracted with corresponding values in the initial frame of the sequence. Multi-class AdaBoost or support vector machine, finally, were applied to classify facial expressions. Regarding hybrid features, Du et al. [196] (2014) extracted both appearance features by Gabor filters and geometric features as the distribution of each pair of fiducials. They then alternatively used the Nearest-mean classifier and Kernel subclass discriminant analysis to classify 6 basic expressions, one neutral expression, and 22 compound expressions. Similarly, Benitez-Quiroz et al. [206] (2016) combined Euclidean distances and angles within normalized facial landmarks, and the facial landmarks after the Gabor filter to recognize a total of 23 expressions using kernel subclass discriminant analysis. **In general, FER using conventional machine learning algorithms demand low computational power and memory. Thus, it can be used for real-time embedded systems** [257]**. However, these methods require hand-crafted feature extraction, which cannot be jointed for optimizing and improving performance** [258], [259]**.**

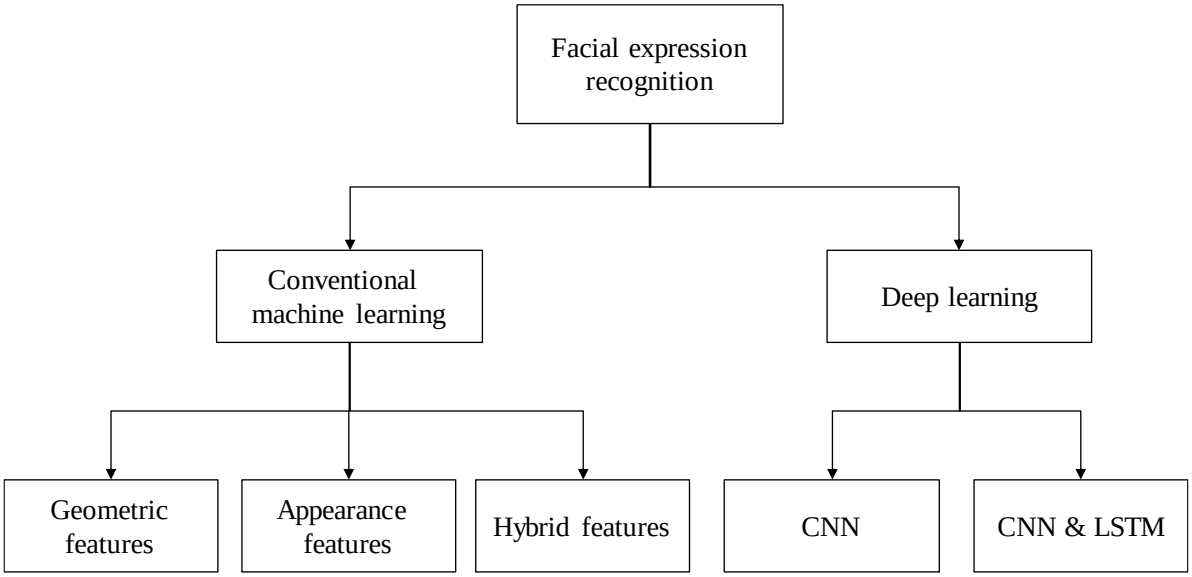


Figure 17. Facial expression recognition conducted on 2D image dataset

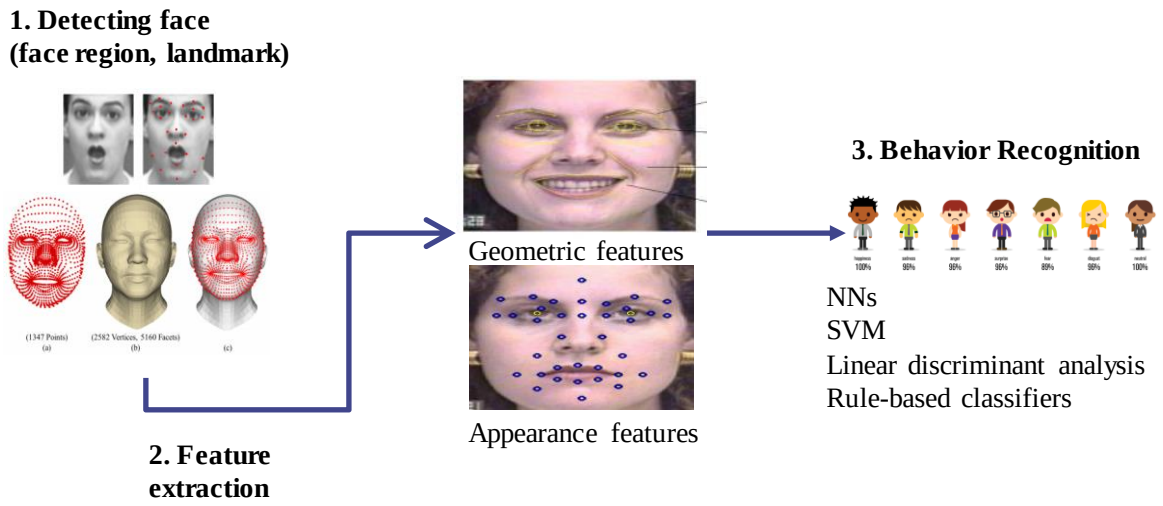


Figure 18. Produce for conduction facial expression recognition using conventional machine learning algorithms.

Deep learning, on the other hand, has a breakthrough and is currently successfully applied in computer vision, especially for facial expression recognition based on 2D image processing (Figure 19). For example, Breuer and Kimmel [260] (2017) trained a CNN model to extract learned features and infer facial expressions with 8 expressions and 50 AUs. Jung et al. [261] (2015) extracted both temporal appearance features and temporal geometry features by training two different CNN models. They then combined two models for recognizing 6 expressions and a neutral expression. In another research, Zhao et al. [262] (2016) proposed an end-to-end trainable network namely deep region and multi-label learning (DRML). The model can induce important facial areas and extract the face's structural information.

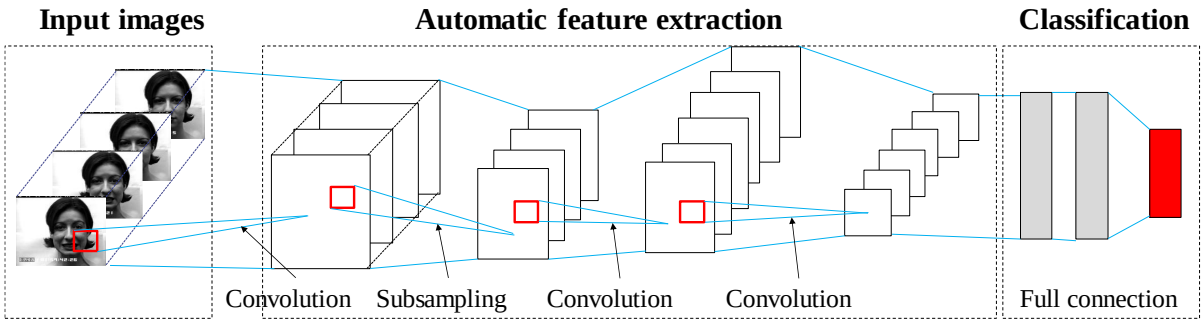


Figure 19. Produce for conduction facial expression recognition using deep learning algorithms [7].

Several studies made use of different deep learning models by combining convolution neural network (CNN) and long-short term memory (LSTM) for classification (Figure 20). An and Liu [263] (2020) combined CNN and LSTM models to extract the feature information from the 2D image dynamic expression face sequences. Firstly, context information was extracted from the input image by performing the parameter adaptive

initialization CNN. Secondly, the main loop RNN and LSTM information memory generate the feature vector from the context information. Finally, facial expression was recognized using the support vector machine (SVM) method. Kumar et al. [264] (2021) proposed a method for facial expression recognition using multi-view representation, which is based on a multi-level uncorrelated discriminative shared Gaussian process model. Li et al. [265] (2020) proposed a new approach for preprocessing data such as removing useless regions to crop the face and rotating the face image to expand the database. A simple CNN consists of two convolution layers with 32 and 64 kernels, two sub-sampling layers to reduce the image size, and one output layer to recognize facial expressions. Ebrahimi Kahou et al. [266] (2015) propagated information about the face in a sequence of facial images with a hybrid RNN-CNN architecture. The model can perform well for recognizing 6 basic expressions and a neutral expression. Kim et al. [267] (2019) also combined CNN and LSTM models. The CNN model learns spatial properties of representative expression-states, while the LSTM model learns temporal properties of the spatial feature representation. Similarly, Chu et al. [268] (2017) also combined CNN and LSTM models for extracting spatial representations and temporal dependencies for detecting multi-level facial AUs for the input video sequences. Hasani and Mahoor [269] (2017), on the other hand, enhanced the 3D Inception-ResNet by adding a LSTM unit for the extraction of spatial and temporal relations between sequence frames in the input video.

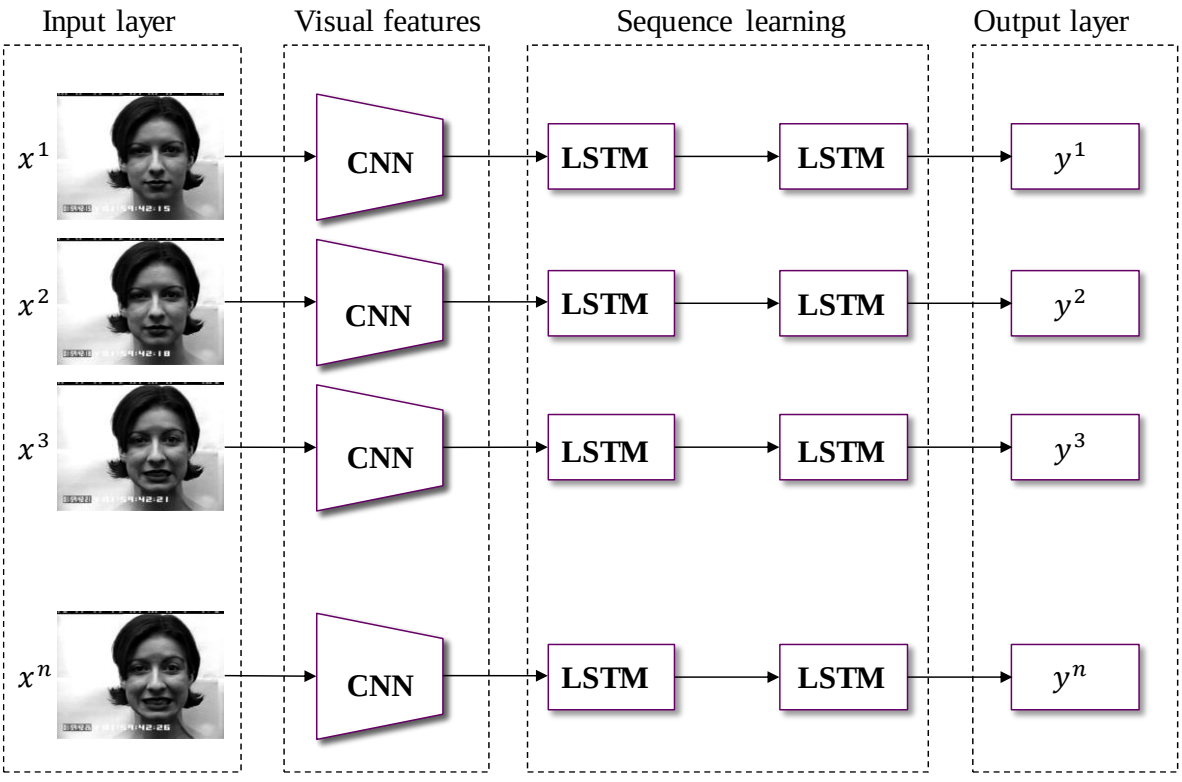


Figure 20. A general framework of a hybrid CNN-LSTM network for facial expression recognition. This hybrid model usually uses CNN for extracting spatial features and combining with LSTM (RNN) for handling temporal features involving sequential images [7].

Generally speaking, deep learning algorithms, unlike conventional machine learning algorithms, estimate features and classify facial expressions using deep neural networks. The methods use deep convolution neural networks to extract optimal features straight from input data. It, however, demands a large amount of data and high computational devices to satisfy deep neural networks.

Various research studies have advanced the accuracy of recognition tasks. However, the recognition performance based on 2D images remains challenging when processing expressions with large variations in different poses and lighting conditions. This task is particularly challenging when performing feature engineering, which is time-consuming and subjective. In fact, 3D information such as the 3D point cloud with an end-to-end deep learning algorithm should be used to deal with the recognition degradation.

2.2.5. 3D facial expression recognition based on 3D face scans

2.2.5.1. Conventional machine learning for 3D facial expression recognition

Feature-based and model-based algorithm

FER on 3D facial databases can be divided into two different types namely feature-based methods and model-based methods [270]. The *feature-based method* extracts feature vectors from surface geometric information of the face. These include spatial relations of interesting points, curvature, gradient, and local shape straight from input 3D face data. These feature vectors then are fed into conventional machine learning algorithms for FER. We can mention support vector machine [271]–[273], hidden Markov model [274]–[276], neural network [273], [277], [278], or random forest [279]. Mohamed Daoudi et al. [280] (2013) reported a fully automatic solution for facial expression recognition from sequences of 3D face scans (BU-4DFE database) that is identity-independent. Firstly, the 3D mesh, frame by frame, was aligned and cropped. Secondly, relevant features from 3D deformations were optimized by Linear Discriminant Analysis (LDA). Hidden Markov Models (HMMs), then, were used for classifying facial expressions based on temporal variations of optimized features.

The *model-based method*, on the other hand, computes coefficients of shape deformation between the input face and the generic face presented by neutral expression. These coefficients are then used as the feature vector. For example, Zhao et al. [281] (2010) tracked the displacements of a set of manually labeling landmarks by fitting statistical facial feature models (SFAM) on the face. These displacements were fed to a Bayesian belief network (BBN) to recognize 6 basic expressions applied to the BU-3DFE database. Hui Chen et al. [282] (2015) reconstructed a 3D face model from a 2D image. A random forest algorithm was then applied for fast facial model-based expression recognition in real-time. The 3D face reconstruction process can handle variations in head poses such as large rotation and fast movements, as well as partial facial occlusions in a sequence of 2D images. Fabiano and Canavan [283] (2018), besides, made use of a statistical shape model to combine it with a local shape index-based feature. It was then applied to 3D facial expression recognition for

both spontaneous and non-spontaneous data (BP4D database). Ocegueda et al. [284] (2011) fitted the 3D facial meshes into an Annotated Face Model (AFM) and estimated Expressive Maps using geometric features such as normal, local curvature, and vertex. These features were then used for classifying expressions.

In general, feature-based methods are mainly based on highly accurate facial landmarks, which then analyzed the configuration and topology of the surrounding regions. Localization of facial landmarks is still an open problem. Model-based methods are, on the other hand, often associated with fitting a generic facial model, which can suffer from big facial deformations (e.g. smiling, widely open mouth).

Some approaches have successfully explored 3D faces for expression recognition tasks by *projecting 3D faces onto the 2D image plane*. Precisely, several techniques project 3D data into the 2D image plane and handle expression recognition based on features extracted from the 2D image plane [285]–[288]. In particular, Savran and Sankur [285] (2017) developed a non-rigid registration method based on projecting 3D faces to the 2D image plane. The method was able to extract two features for expression detectability. 1) The permanent face structures are lower within-class variance. And 2) expression face structures are higher between-class variance. Another research by Savran et al. [287] (2012) expressed the action unit intensity estimation based on a nonlinear scale of 5 grades using both 2D and 3D information. Firstly, facial features are extracted by applying Gabor wavelets on 2D luminance images. 3D surface geometry images (local shape feature) such as mean curvature, Gaussian curvature, shape index, and curvedness are mapped onto the 2D surface. These local shape features are estimated from the maximal and minimal principal curvatures. Secondly, a support vector machine regression was implemented on the coefficients of Gabor wavelets to find the score of action unit intensity.

Some other approaches *find a new presentation* of the 3D face in terms of curvature descriptor [289], Scale-invariant Feature Transform (SIFT) feature [290], and spatial distances and displacements between facial landmarks [291], [292]. In particular, Alyuz et al. [289] (2008) focused on using curvature descriptors, which measure the bending degree of a surface as the feature vector for the classification task. Berretti et al. [290] (2011) computed SIFT descriptors for a set of facial key-points for the BU-3DFED dataset and used these descriptors as feature vectors for recognition tasks using a support vector machine model. Berretti et al. [291] (2013) also proposed a method for automatically recognizing facial expressions for dynamic 3D face scan sequences. They detected a set of 3D facial landmarks. A new presentation of the face was constructed comprising of mutual spatial distances between these facial landmarks, and 3D shape context as the facial landmark descriptors of the neighbor points. Zarbakhsh and Demirel [292] (2020) proposed a time series analysis method to process the dynamic 3D facial expression. Firstly, the topological features such as distance, displacements, and angles are extracted from 3D facial landmarks. Secondly, the neighborhood component feature selection was applied to reduce the high dimensional feature space. Thirdly, the temporal representation of geometric deformation values was used to construct a multimodal time series model. At last, recognition was performed by adaptive

cost dynamic time warping classification. Lemaire et al. [293] (2013) described 3D facial surfaces as a set of 2D maps that might make use of existing 2D image processing techniques. 3D face models were first preprocessed by aligning into a frontal pose and cropping the face into the center region. Differential Mean Curvature Maps (DMCMs) were then used to extract Representation Maps including curvature at different scales, which can generate different 2D representation images. After that, these DMCMs were processed using the Histogram of Oriented Gradients (HOG) to extract global descriptors. And finally, a multi-class support vector machine was applied for classifying facial expressions.

Multi-models using 2D and 3D facial data

Different features such as facial shapes, curvature, facial landmarks, and texture information could be extracted independently from both 2D images and 3D face scans and emerge to obtain state-of-the-art performance for recognizing facial expressions [294]. In particular, Hayat and Bennamoun [295] (2013) separately learned features from 2D images and 3D face scans and use SVM to classify facial expressions. Jan and Meng [296] (2015) emerged geometric features from 3D databases and texture features from 2D images for FER.

2.2.5.2. Deep learning for 3D facial expression recognition

Deep learning has recently been applied for recognizing 3D facial expressions. For example, Oyedotun et al. [297, p.] (2017) learned discriminative features using a deep CNN model. They fused 2D images and depth map latent representations. The method was able to differentiate 6 expressions for the BU-3DFE database. Yang and Yin [298] (2017) also tested 3D facial expression recognition using CNN and facial landmarks for two databases BU-3DFE and BU-4DFE. Li et al. [299] (2015) extracted the first deep representation such as the geometry map, the normalized curvature map, and normal maps from 3D faces. They also extracted the texture map from 2D photometric using a pre-trained deep CNN. These representations were then used to recognize expression by SVM. Chen et al. [300] (2018) proposed a Fast and Light Manifold CNN model, which can enhance geometric representation.

The 3D point cloud is the representation of the surface of a 3D object in space. A simple format of the point cloud is a set of data points in terms of XYZ coordinates [301]. Thus, point cloud resolution relates to the density of the number of points within the 3D object surface. Geometric data such as 3D facial point clouds are not a regular format. In general, previous studies often project this data into 2D images or find a new presentation of the 3D points, and then recognize the facial based on these new presentations.

To summarize, most of these researches focus on face registration and then extracting the feature for facial expression recognition. Or they represent the 3D face in terms of a new presentation such as curvature descriptors, SIFT feature, spatial distances, or displacements. This, however, transforms the data and often reduces the depth information on the face. To the best of our knowledge, there is no research study using directly the 3D point cloud for facial expression recognition. Therefore, we propose to use a new class of deep learning called geometric deep learning model, called PointNet++ [302] based on PointNet [303] to directly process 3D geometry such as 3D facial point cloud for facial expression recognition. More details on this part will be presented in Chapter 4.

2.3. Facial symmetry analysis

Facial symmetry indicates the balance and equality of facial structure in terms of shape, size, location, and arrangement of left and right components on the sagittal plane [304]. On the opposite, the asymmetric face shows the bilateral difference between the two sides. Facial symmetry is a crucial factor in recognizing the impression of beauty and attraction [305]. Indeed, people with attractive faces are more likely to get more chances in social activities such as dating [306], [307], getting hired at prestigious occupations [306], or even getting benefits from social advantages in daily life [308]. Moreover, facial palsy patients and patients with facial transplantation have facial asymmetry, which leads to unnatural facial expressions [31], [107]. In fact, facial expressions are important due to contributing 55% to the emotional interpretation of the conversations [309]. Thus, being unconfident when performing facial expressions could lead to a loss of confidence in face-to-face conversation and reduce the patient's confidence resulting in self-isolation and psychological disorders [18], [107]. Consequently, facial symmetry analysis is important to support the correct assessment and diagnosis of facial palsy and facial transplantation patients. This could help the doctor to design efficient individual rehabilitation programs [21], [33].

Grading and analyzing the symmetry of the face could be categorized into traditional and computer-aided methods. *Traditional methods* were subjectively conducted such as House-Brackmann system [35] (1985), Burres-Fisch system [310] (1990), and Sunnybrook system [34] (2010). This was mainly based on measuring Euclidian distances between pre-defined facial landmarks at resting or facial expression positions. *Computer-aided systems* can provide quantitative and objective measurements, it has mainly based on 2D image processing. For example, the Neely-Cheung rapid grading system analyzed the face by entering the data measured by hand into the system [311], [312] (1999). Pourmomeny et al. [313] (2011) proposed a system to grade the face to measure the face during facial movements using Photoshop software. SmartEye Pro-MME (2011), which tracks lip movements, requires manually picking facial landmarks on the image [314]. Another system could measure the symmetry of the face at the resting position using Kinect v2 sensor [315]. In general, previous studies have two limitations. The first limitation is that it required hand-crafted feature engineering techniques (e.g. facial landmark detection). The second limitation is that it was reduced performance when processing images with large variations in different expressions, pose, and lighting conditions. Thus, to deal with these challenges, 3D information such as 3D point cloud data using end-to-end solutions to find a novel shape descriptor of the data should be used for facial symmetry analysis.

The breakthrough of novel sensing devices leads to the availability of non-Euclidean data such as 3d point clouds, graphs, and meshes [316]. And 3D point cloud databases are the regular choice for capturing the shape of subjects as they are flexible and efficient [317].

The shape descriptors are a new presentation of the object and should capture the distinctive properties of the object's geometric structure [318]. Recently years, many studies have been proposed to have effective feature extraction of 3D point cloud descriptors for different computer vision tasks for classification and segmentation. Several examples of

descriptors are shape distribution [319], 3D Zernike descriptor [320], Fourier descriptor [321], eigenvalue descriptor [322], spin images [323], mesh HOG [324], and shape context [325]. In general, in terms of the way to extract the feature, 3D shape descriptors can be divided into two different categories: hand-crafted based descriptors and deep learning-based descriptors [326]. However, having discriminative and robust novel shape descriptors remains a challenge and is worth being investigated due to its nature of unstructured and unorder points.

To sum up, facial symmetry is an important indicator for analyzing facial palsy. Analyzing facial symmetry can support the correct assessment of involved patients. Previous studies have mainly been based on either subjective methods conducted by doctors or automatic systems. These methods require hand-crafted feature engineering (e.g. facial landmark detection), which is time-consuming and challenging due to image variations. We proposed to use a novel point descriptor from 3D point cloud data for analyzing the face in terms of facial symmetry. A novel type of learned descriptor is estimated from the state-of-the-are geometric deep learning model called PointNet++ [302]. The descriptors were then applied principal component analysis to reduce the dimension [326]. The final descriptors were then used to analyze the facial symmetry and in the meantime the difference between the face of Caucasians and Asians. Another descriptor was also investigated called the hand-crafted method based on the PointPCA model [385]. More details will be presented in Chapter 5.

2.4. Facial motion learning by coupling between reinforcement learning and finite element model of the face

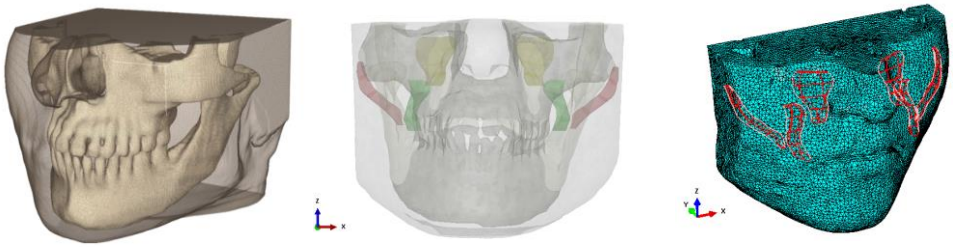
2.4.1. Introduction

Facial palsy patients or patients with facial transplantation have abnormal facial movement patterns due to altered facial muscle functions and nerve damage. This leads to abnormal motion control for different movements such as eating, speaking, or facial expressions [11]–[13], [309]. Moreover, involved patients also suffer asymmetric face effects. This effect indicates the imbalance and inequality of facial structure in terms of shape, size, location, and arrangement of left and right components on the sagittal plane [304]. In fact, the recovery to a symmetric face with balanced functionalities requires a complex rehabilitation process. This process requires patients must practice rehabilitation exercises. Long-term plastic changes in the brain can be induced by repetitive sensory and motor exercises during rehabilitation. In fact, these repetitive rehabilitation exercises could assist the brain, facial muscles, and facial nerve systems coordinate to reroute the electrical signals from the brain to muscles that were interrupted for involved patients [109], [327]. In addition, knowing which muscle is responsible for what movements could help clinicians identify straightforward muscles that should be targeted for surgery [328]. Thus, understanding of facial motion mechanism remains a scientific and clinical challenge to help the involved patients to recover symmetrical movements and normal facial expressions. It is important to note that current facial rehabilitation has mainly been based on a mirror approach to monitor the visual qualitative feedback from the rehabilitation exercise. More precisely, patients watch their distorted features in the mirror as a reference to teach themselves the right expressions during rehabilitation exercises. This strategy is ineffective and subjective without any feedback. Moreover, the current rehabilitation process is limited by a lack of patient-specific knowledge about muscles driving facial motions. Therefore, understanding facial motion mechanisms, muscle activation, and coordination are clearly fundamental. To provide quantitative and objective information on the facial motion during the rehabilitation exercise, computer-based systems that automatically recognize action units (AUs) defined by the Facial Action Coding System (FACS) have been developed [40]. Such complex systems can provide an objective guideline for monitoring the facial rehabilitation process, which is long-term, inconvenient, and sometimes ineffective [33], [329], [330].

2.4.2. Biomechanical model for learning motion patterns

In addition, for investigating muscles driving facial motion problems, biomechanical models are recommended. Because they can be customized to reflect the true anatomy and pathological anatomical deformations as well as imitate physical processes [331]. In particular, these biomechanical models also provide an effective and powerful tool for assessing the structure and functionalities of the human anatomy. This is of help in developing a scientific foundation for treatment planning. Modeling is especially important when mechanical parameters are difficult or impossible to quantify with the available

technologies. In fact, physics-based facial models using finite element methods have been intensively developed to explore the role of facial muscle excitation, contraction, and coordination during facial motion (Figure 21) [15], [16], [122], [328], [332]–[337].



(a) (b) (c)

Figure 21. A physical-based model of the face with the skull in soft tissue (a), a physical-based model of the face with reconstructed muscles (b), and a physical-based model of the face using finite element modeling (c) (Fan 2016) [338]

In the physical-based facial model system, muscle excitation is an unknown variable. Muscle excitation represents the neural control process. Muscle helps to contract the face tissues and move the skull to perform facial expressions and movements. Physical-based model simulations provide a detailed view of the muscle contraction mechanism and its effect on facial motion. However, the physics-based approach is descriptive with a priori knew input information such as muscle properties. **Moreover, which muscles and what value of muscle excitations for performing the desired movement for facial rehabilitation is still an open and longstanding research question**. It is almost practically hard or impossible to directly measure muscle activations from living subjects due to safety and accessibility limitations [332]. In fact, muscle force can be manually measured in extreme conditions. For example, it can be measured in the operation room by placing a force transducer on a tendon that collects information and was removed before finishing the surgery (e.g. finger flexor tendon forces measurement [339]). This method, however, mostly be used in research laboratory environments. Otherwise, measuring muscle force could rely on the use of skeletal movements as well as ground reaction forces. This method, nonetheless, could not provide affection for any single muscle. **Diverse numerical techniques have been proposed for estimating muscle excitation such as inverse dynamics, forward-dynamics tracking simulation, and optimal control strategies** [328]. Precisely, a technique namely inverse dynamic relied on computational modeling of the dynamics of connected multi-bodies. This approach is based on solving a static optimization problem to discover the most likely collection of muscle excitations that will result in desired movements of the biomechanical model for each timestep [340]. The forward-dynamics assisted tracking, on the other hand, feds an initial set of muscle excitations into forward dynamics equations of the biomechanical model. It then iteratively updates the muscle excitations based on cost function comparing the computed result and experimental kinematics [328]. **However, the use of these approaches depends strongly on the a priori definition of input data, model properties, and the targeted motion**. Another significant limitation of any inverse dynamic solution is that it

depends on the availability of movement patterns as the input. This information is not always simple to collect from living subjects, especially in the case of patients with facial palsy or facial transplantation. Thus, this approach has a limited predictive capacity to explore a larger parameter space to find emerging properties during dynamic movements of the face.

2.4.3. Reinforcement learning for inverse dynamics motion learning

In spite of the increasing availability of massive databases and computational models, artificial intelligence has rapidly grown [341]. One of the promising solutions in the control field is reinforcement learning with tremendous theoretical and practical achievements in robotics control [342], gamming [343], autonomous driving [344], computer vision [345], and healthcare [341], [346], [347]. In particular, the question of the use of this learning strategy in the healthcare domain to tackle real-world applications has recently been raised [341], [346], [347]. Reinforcement learning distinguishes itself from other types of machine learning in several perspectives. The agent collects data through interactions with the environment and uses that data to train the agent itself. This dependence leads to variation outcomes from one run to another. Recently, reinforcement learning strategy has been coupled with rigid multi-bodies dynamics to explore the motion of the lower limbs during walking and age-related falls in our team [348]. In another research, Izawa et al. [349] (2004) proposed to use an actor-critic reinforcement learning algorithm to learn to control a biological arm model. Broad [350] (2011) learned for achieving desired arm movements using reinforcement learning. Abdi et al. [351] (2020) estimated muscle excitation for biomechanical simulation based on a deep reinforcement learning algorithm. Thus, this learning strategy opens new avenues to explore human system motion and novel emerging properties without any a priori motion data as well as much knowledge of the biomechanical systems.

To sum up, the facial motion learning capacity will be explored by the coupling between reinforcement learning and finite element modeling. The main objective is to provide, for the first time, the modeling workflow for this complex coupling and then to evaluate different learning strategies to establish motion patterns of the face during facial expression motions. Our novel solution will explore the patient-specific facial motions without a priori data from the patient and then provides a set of facial muscle activation and coordination patterns for a specific rehabilitation-oriented movement (e.g. symmetry or smile). More details will be presented in Chapter 6.

2.5. Conclusion

This chapter summarized the literature review about 3D face reconstructions from 2D images, facial expression recognition, facial symmetry analysis, and facial motion learning based on biomechanical models. Throughout this review, several challenges for clinical applications have been pointed out. The first challenge is the lack of building 3D information from images or depending strongly on the selected depth cameras. The second challenge is the lack of analyzing the face in terms of expression recognition and symmetry analysis. The last challenge is the limitation of the predictive capacity of the facial motion patterns with emerging biomechanical properties.

Chapter 3:

3D Face Reconstruction from a Single Image using Different Deep Learning Approaches for Facial Palsy Patients (2D_3D)

Analyzing the face of facial palsy or facial transplantation patients helps doctors and patients to assess and improve the rehabilitation process. Directly capturing a 3D face using a medical imaging device such as **MRI** is high accuracy but that is expensive and time-consuming. The use of **depth cameras** like **Kinect** can lead to a reasonable accuracy of the 3D face model while keeping the cheap cost. But it has to be replaced in the future application due to stopped production. **3D face reconstruction from a 2D image** seems like a solution. A huge effort has been invested in estimating the subject-specific 3D surface of the face from 2D images under uncontrolled and uncalibrated conditions. Existing approaches could output good accuracy levels for 3D face reconstruction and are able to reconstruct 3D subject-specific face. However, the 3D face shape reconstruction of facial palsy patients is still a challenge and this has not been investigated. In this chapter, we applied several well-established methods to reconstruct 3D faces of 4 subjects including 2 healthy subjects and 2 facial palsy patients in neutral and mimic postures from one single 2D image from Kinect v2. The methodology can also be applied to 2D images captured from any kind of device. This part corresponds with task number 1 (Figure 22).

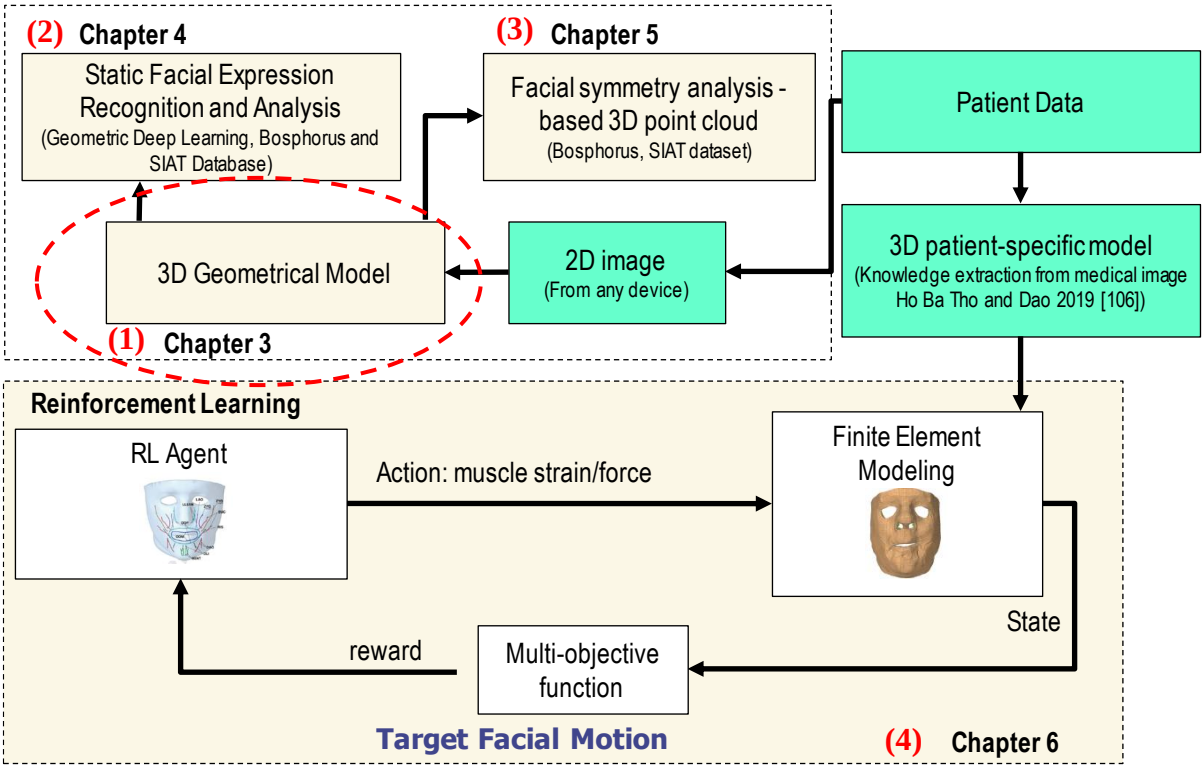


Figure 22. Thesis framework: the part of 3D face reconstruction aims to generate 3D geometrical model of the face

3.1. Materials and Methods

In this part, we proposed to used three methods for reconstructing the subject-specific face model from a single 2D image. The first method was based on fitting a 3DMM to the input 2D image. The second and third methods were based on regressing parameters of 3DMMs using a pre-trained ResNet-50 model.

3.1.1. Materials

In order to reconstruct the 3D face from a 2D image, we used a dataset of 2 healthy subjects (one male and one female) and 2 facial palsy patients (two females) collected from CHU Amiens (France). Each healthy subject or patient signed an informed consent agreement before the data acquisition process. The protocol was approved by the local ethics committee (no2011-A00532-39). The subject performed several trials with neutral position and facial mimic positions such as smile, [e], and [u] pronunciation. Our developed Kinect-based computer vision system [109] was used to capture the high density (HD) point clouds of the face as well as the RGB image from Kinect sensors. The images were captured where each subject was positioned in front of the camera. The RGB image was used for 3D shape reconstruction with the deep learning models. The HD point cloud was used to reconstruct the 3D shape for validation purposes. Moreover, 3D face scans using MRI were also available for validation purposes. Two other datasets were collected in order to test more patients with facial palsy. The first dataset of 8 patients was collected in an unconstrained condition [352]. The second dataset of 12 patients are obtained from the Service Chirurgie Maxillo-Faciale CHU Amiens (Prof. Stéphanie DAKPE, Dr. Emilien COLIN) from a pilot study « Etude pilote d'évaluation quantitative de l'attention portée aux visages présentant une paralysie faciale par oculométrie (eye-tracking) » with clinical trial registered (ClinicalTrials.gov Identifier: NCT04886245 - Code promoteur CHU Amiens-Picardie: PI2019_843_0089 - Numéro ID-RCB: 2019-A02958-49).

3.1.2. Methodology

3.1.2.1. Method 1: Fitting a 3D Morphable model

The information processing pipeline of the 3D morphable modeling approach [129] to reconstruct the 3D face shape from a single image is illustrated in Figure 23. Firstly, a set of 2D facial landmarks are detected from the input image by existing face detectors [353], [354]. Secondly, the scaled orthographic projection projects another set of landmarks from the 3D model to get 2D points in the image plane corresponding with those points obtained from the 2D image. This step results in an equation that parameterizes the pose and shape parameter. In the next step, a cost function is built to minimize the error between the 2D facial landmarks from the 3D model and 2D facial landmarks from the 2D facial image.

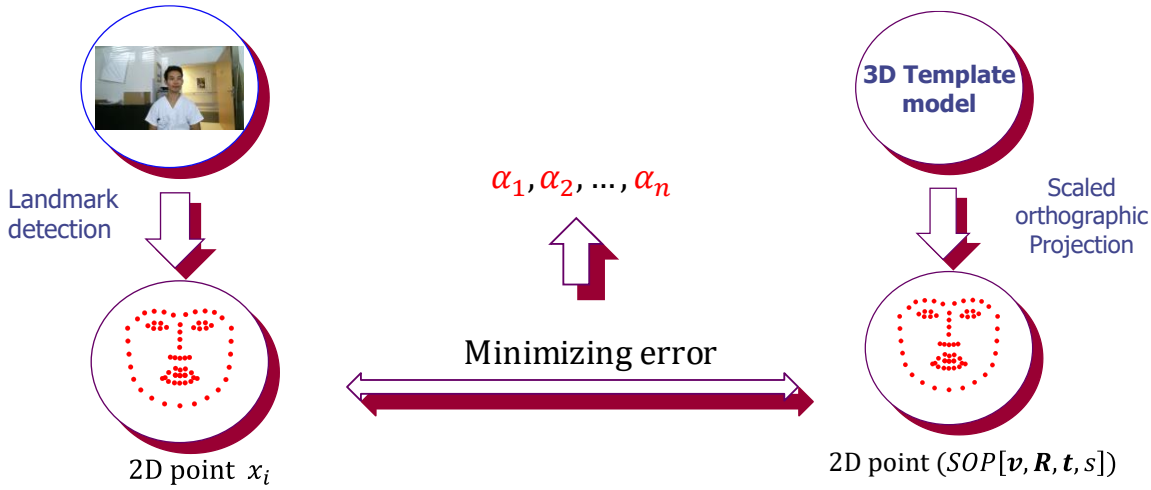


Figure 23. Pipeline to estimate the shape parameters of the 3DMM

3.1.2.1.1. 3D Basel Morphable Model

In the present study, the Basel 3D Morphable model (3DMM) was used [133]. This model was built from a set of 3D faces from a scan of 100 females and 100 males by presenting the face model in terms of trained vector spaces as shape vector spaces. Each face is parameterized in the form of angular meshes with 53490 vertices. The $\boldsymbol{s} = (x_1, y_1, z_1, \ldots, x_m, y_m, z_m)^T$ is the shape vector for $m = 53490$ vertices. Each vector is in $53490 \times 3 = 106470$ dimension.

In the next step, all shape vectors of all 200 subjects were concatenated to obtain the matrix of shape $\boldsymbol{S}$ (106470×200 dimensional matrix). The Principal Component Analysis (PCA) was utilized to decompose the shape matrix resulting in a set of linear combinations of shape bases and texture bases. The PCA is usually a technique for reducing the dimension of the high dimensional data and still remains the largest information. The constitutive equation of this approach is given in the following equation:

$$\boldsymbol{S} = \boldsymbol{S}_0 + \boldsymbol{S}_i \alpha_i$$

where $\boldsymbol{S}$ (106470×200 dimension matrix) is the shape matrix of 200 subjects, $\boldsymbol{S}_0$ (106470×200 dimensional matrix) is the mean shape matrix with a mean shape vector at each column. $\boldsymbol{S}_i$(106470×106470 dimensional matrix) is the principal component of the eigenvector of the covariance matrix from the shape matrix and α_i (106470×200 dimensional matrix) is the eigenvalue, which stands for the coefficients of the shape. The number of the $\boldsymbol{S}_i$ column and the α_i row can be reduced by choosing the large value of eigenvalue and dropping out the less value of eigenvalue.

In the PCA decomposition, the mean shape and the shape bases are shared for every specific individual. This means that the mean shape $\boldsymbol{s}_0$ and the shape bases ($\boldsymbol{S}_i$) are the same for every subject in the training set, those can be assumed as the population and can be used

for other subjects different from the subject in the training set. Therefore, from a given facial image, the 3D face can be reconstructed by finding the coefficients of the specific shape.

3.1.2.1.2. Model fitting

Facial landmark detection

Based on the input image, facial landmarks were detected using the pre-trained facial landmark detector (dlib library) for the iBUG300-W database [353] from "300 Faces In-the-Wild Challenge" for automatic facial landmark detection. The method detects 68 facial landmarks using the Active Orientation Models, which is a variant of Active Appearance Models [355].

Pose from Scaled Orthographic Projection

The rotation matrix and translation vector of the face were used to transform a 3D face from the space coordinate system into the camera system. The scaled orthographic projection assumes that the depths from every point in the face to the camera are not various from one another, therefore, the mean depth of the face can be the same for every point on the face. The projection of the 3D face to the image plane and then can be estimated using rotation $\boldsymbol{R} \in \mathbb{R}^{3\times3}$, translation $\boldsymbol{t} \in \mathbb{R}^2$, and the scale factor $s \in \mathbb{R}$. This is expressed in the following equation:

$$\boldsymbol{SOP}[\boldsymbol{f},\ \boldsymbol{R},\boldsymbol{t},s] = s\begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 0\end{bmatrix}\boldsymbol{Rf} + s\boldsymbol{t}$$

where, the $\boldsymbol{SOP}[\boldsymbol{f},\ \boldsymbol{R},\boldsymbol{t},s]$ is the 2D points of the 3D face in the image plane by scaled orthographic projection; $\boldsymbol{f}$ represents the 3D facial points.

Additionally, the $\boldsymbol{f}$ in the projection equation can be expressed using the shape equation as follows:

$$\boldsymbol{f} = \boldsymbol{f}_0 + \boldsymbol{f}_i\alpha_i$$

where $\boldsymbol{f}$ is several points in the 3D face, $\boldsymbol{f}_0$ is the corresponding points of the mean shape face, $\boldsymbol{f}_i$ is the corresponding shape bases and α_i is the shape coefficients that can be used to reconstruct the 3D face.

Fitting correspondences

Optimizing the difference between 2D facial landmarks detected from the input image and corresponding 2D facial landmarks projected from the 3D model could result in the pose parameter including rotation, translation, and scale factor along with the shape parameter (shape coefficients) as follows:

$$\boldsymbol{E}[\boldsymbol{\alpha},\ \boldsymbol{R},\boldsymbol{t},s] = \frac{1}{\boldsymbol{L}}s\sum_{i=1}^{L}\|x_i - \boldsymbol{SOP}[\boldsymbol{v},\ \boldsymbol{R},\boldsymbol{t},s]\|$$

This problem can be solved by the iterative algorithm POSIT (POS with Iteration) [356]. The solution is able to estimate the rotation $\boldsymbol{R}$, translation $\boldsymbol{t}$, and scale factor s of the input face and the shape parameter α_i for reconstructing the 3D face of the specific image.

3.1.2.2. Method 2: DECA

The second method reconstructs the 3D face using an approach of Feng et al. [357] (2021), in which the coefficients of the FLAME model [358] were learned from the pre-trained model ResNet50 [175].

3.1.2.2.1. The principle

The geometry shape method used an established 3D statistical head model namely FLAME [358], which can generate the face with different shapes, expressions, and poses. The model is a linear combination of identity $\boldsymbol{\beta} \in \mathbb{R}^{|\beta|}$, expression $\boldsymbol{\psi} \in \mathbb{R}^{|\psi|}$ with linear blend skin, and pose $\boldsymbol{\theta} \in \mathbb{R}^{3k+3}$ ($k = 4$ includes the neck, jaw, and two eyeballs). The FLAME model is defined as follows:

$$M(\boldsymbol{\beta}, \boldsymbol{\psi}, \boldsymbol{\theta}) = W(T_P(\boldsymbol{\beta}, \boldsymbol{\psi}, \boldsymbol{\theta}), \mathbf{J}(\boldsymbol{\beta}), \boldsymbol{\theta}, \boldsymbol{\mathcal{W}})$$

$$T_P(\boldsymbol{\beta}, \boldsymbol{\psi}, \boldsymbol{\theta}) = \mathbf{T} + B_S(\boldsymbol{\beta}, \boldsymbol{S}) + B_P(\boldsymbol{\theta}, \boldsymbol{\mathcal{P}}) + B_E(\boldsymbol{\psi}, \boldsymbol{\mathcal{E}})$$

where $W(\mathbf{T}, \mathbf{J}, \theta, \boldsymbol{\mathcal{W}})$ is the blend skinning function rotating a set of vertices in $\mathbf{T} \in \mathbb{R}^{3n}$ around joints $\mathbf{J} \in \mathbb{R}^{3k}$, which smoothed by the blend weights $\boldsymbol{\mathcal{W}} \in \mathbb{R}^{k \times n}$.

Appearance model was converted from the Basel Face Model and generated a UV albedo map $A(\boldsymbol{\alpha}) \in \mathbb{R}^{d \times d \times 3}$, where albedo parameter $\boldsymbol{\alpha} \in \mathbb{R}^{|\boldsymbol{\alpha}|}$.

Camera model aims to project 3D vertices onto the image plane $v = s\Pi(M_t) + t$, where $M_t \in \mathbb{R}^3$ is a vertex in M, $\Pi \in \mathbb{R}^{2\times3}$ is the orthographic projection matrix from 3D to 2D, and $s \in \mathbb{R}$, and $t \in \mathbb{R}^2$ represent isotropic scale and 2D translation respectively.

Illumination model finds the shaded face image based on Spherical Harmonics [359]. And texture rendering is based on the geometry parameters $M(\boldsymbol{\beta}, \boldsymbol{\psi}, \boldsymbol{\theta})$, albedo and camera information.

3.1.2.2.2. Model learning

Reconstructing the face of the patients based on two steps: coarse reconstruction and detail reconstruction.

A coarse reconstruction was performed by training an encoder E_c consisting of ResNet50, which minimizes the variation between the input image I and the synthesis image I_r, which is synthesized and generated by decoding the latent code of the encoded input image. The latent code contains a total of 236 parameters of the face model such as geometric information (100 shape parameters of $\boldsymbol{\beta}$, 50 expression parameters of $\boldsymbol{\psi}$, and pose parameters $\boldsymbol{\theta}$), 50 parameters of appearance information $\boldsymbol{\alpha}$, camera and lighting conditions.

The loss function for the E_c network computes the differences between input image I and the synthesis image I_r and consists of 1) landmark loss ($L_{landmark}$) of 68 2D key points on the face, 2) eye closure loss (L_{eye}) penalizes the relative variation between landmarks on the upper and lower eyelid, 3) photometric loss ($L_{photometric}$) compares between input image I and the synthesis image I_r, 4) identity loss ($L_{identity}$) computes the cosine similarity which presents the fundamental properties of the patient's identity, 5) shape consistency loss (L_{shape}) computes the differences between the shape parameters ($\boldsymbol{\beta}$) from different images of the same patient, and 6) regularization ($L_{regularization}$) for shape, expression, and albedo as follows:

$$L_{coarse} = L_{landmark} + L_{eye} + L_{photometric} + L_{identity} + L_{shape} + L_{regularization}$$

Then **the detail reconstruction** assists to augment the coarse reconstruction with the different details such as wrinkles, and facial expressions using a detailed UV displacement map. The detail reconstruction trains an encoder E_d, which is the same architecture E_c to output 128 latent code $\boldsymbol{\delta}$ relates to the patient-specific details. The loss function for E_d network contains 1) photometric detail loss ($L_{photometric\ detail}$) based on detail displacement map, 2) implicit diversified Markov random field loss (L_{mrf}) [360] relates to geometric details, 3) soft symmetry loss ($L_{symmetry}$) to cope with self-occlusions of face parts, and 4) detail regularization ($L_{regularization\ detail}$) to reduce noise as follows:

$$L_{detail} = L_{photometric\ detail} + L_{mrf} + L_{symmetry} + L_{regularization\ detail}$$

3.1.2.3. Method 3: Deep 3D Face Reconstruction

The third method relates to the deep 3D Face Reconstruction approach of Deng et al. [361] (2020), in which the coefficients of the 3D morphable model of the face were learned from the pre-trained model ResNet50 [175].

3.1.2.3.1. 3D Morphable Model

The shape $\boldsymbol{S}$ and the texture $\boldsymbol{T}$ of the 3DMM were presented as follows:

$$\boldsymbol{S} = \boldsymbol{S}(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \overline{\boldsymbol{S}} + \boldsymbol{B}_{id}\boldsymbol{\alpha} + \boldsymbol{B}_{exp}\boldsymbol{\beta}$$
$$\boldsymbol{T} = \boldsymbol{T}(\boldsymbol{\delta}) = \overline{\boldsymbol{T}} + \boldsymbol{B}_{t}\boldsymbol{\delta}$$

where $\overline{\boldsymbol{S}}$ and $\overline{\boldsymbol{T}}$ are the mean shape and texture of the face model; $\boldsymbol{B}_{id}$, $\boldsymbol{B}_{exp}$, and $\boldsymbol{B}_t$ are the principal component vectors based on PCA presenting for identity, expression, and texture; and respective coefficients vectors $\boldsymbol{\alpha}, \boldsymbol{\beta}$, and $\boldsymbol{\delta}$.

The scene illumination was modeled using Spherical Harmonics coefficients $\gamma_b \in \mathbb{R}^9$. The radiosity of a vertex s_i was computed as $C(\boldsymbol{n}_i, \boldsymbol{t}_i) = \boldsymbol{t}_i \cdot \sum_{b=1}^{B^2} \gamma_b \Phi_b(\boldsymbol{n}_i)$, where $\boldsymbol{n}_i$ and $\boldsymbol{t}_i$ are the surface normal and skin texture of the vertex s_i, Φ_b is Spherical Harmonics basis functions.

The pose $\boldsymbol{p}$ of the face is represented by rotation **R** and translation **t**. All the unknown parameters (e.g. $x = (\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\delta}, \gamma, \boldsymbol{p}) \in \mathbb{R}^{239}$) are the output of the modified RestNet-50 with the last layer including 239 neurons.

3.1.2.3.2. Model learning

The coefficients are the output of the ResNet-50 model as illustrated in Figure 24, which is modified last fully collected layer and was trained by estimating a hybrid-level loss of image-level loss and perception-level loss instead of using ground truth labels.

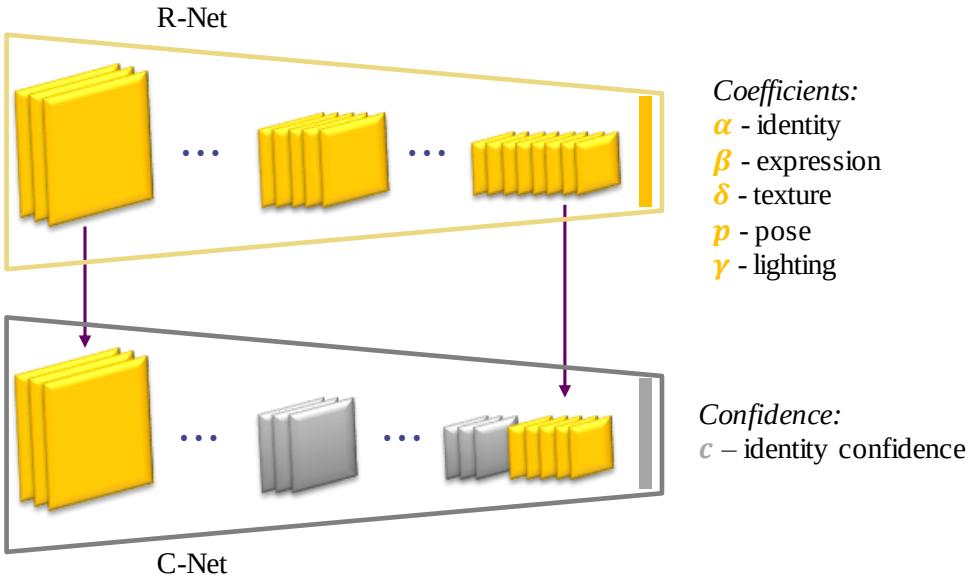


Figure 24. The network architecture for learning the parameters of the face model

Image-level losses integrate photometric loss for each pixel and landmark loss for sparse 2D landmarks detected from the input image. The photometric loss between the raw (I) and the reconstructed (I') images was defined as follows:

$$L_{photo}(x) = \frac{\sum_{i \in \mathcal{M}} \mathrm{A}_i \cdot \|I_i - I_i'(x)\|_2}{\sum_{i \in \mathcal{M}} \mathrm{A}_i}$$

where with each pixel index i, $\mathcal{M}$ denotes re-projected face region, A is skin color, and $\|\cdot\|_2$ is the l_2 norm.

Landmark loss is computed on 68 landmarks $\{\boldsymbol{q}_n\}$ detected from input image [362] and landmarks projected of the reconstructed shape onto the image $\{\boldsymbol{q}_n'\}$ as follows:

$$L_{lan}(x) = \frac{1}{N} \sum_{n=1}^{N} \omega_n \|\boldsymbol{q}_n - \boldsymbol{q}_n')\|^2$$

where ω_n is the landmark weight and set to 20 for the mouth, and nose points, while to 0 for others.

Perception-level loss tackles the local minimum issue for CNN-based reconstruction by extracting deep features from the images of the pre-trained FaceNet model for deep face recognition [49] and uses it to estimate perception loss.

$$L_{per}(x) = 1 - \frac{\langle f(I), f(I'(x))\rangle}{\|f(I)\| \cdot \|f(I'(x)\|}$$

where $f(\cdot)$ represents the deep feature, $\langle \cdot, \cdot \rangle$ is the vector inner product.

Two regularization losses involving coefficients and textures are added to avoid shape and texture degeneration. The coefficients loss invokes the distribution close to the mean face:

$$L_{coef}(x) = \omega_\alpha \|\alpha)\|^2 + \omega_\beta \|\beta)\|^2 + \omega_\gamma \|\delta)\|^2$$

The weights are set $\omega_\alpha = 1.0$, $\omega_\beta = 0.8$, and $\omega_\gamma = 0.0017$. The texture loss is computed by flattening constrain

$$L_{tex}(x) = \sum_{c\in\{r,g,b\}} var\left(\mathrm{T}_{c,\mathcal{R}(x)}\right)$$

where $\mathcal{R}$ is a pre-defined region of the skin at the cheek, nose, and forehead.

3.1.3. Validation versus Kinect-driven and MRI-based reconstructions

The reconstructed outcomes from the above three methods were compared to the 3D shape reconstructed from the Kinect-driven and MRI-based shapes. The 3D Kinect-driven shape was reconstructed by using a computer vision system developed by our team [111]. Specifically, the MRI (magnetic resonance imaging) images were segmented using the semi-automatic method with the 3D Slicer software as shown in Figure 25. 3D shapes were saved in the STL format for further comparison. Hausdorff distance [363] was used to estimate the error of the reconstructed face compared with ground truth data from MRI and Kinect devices.

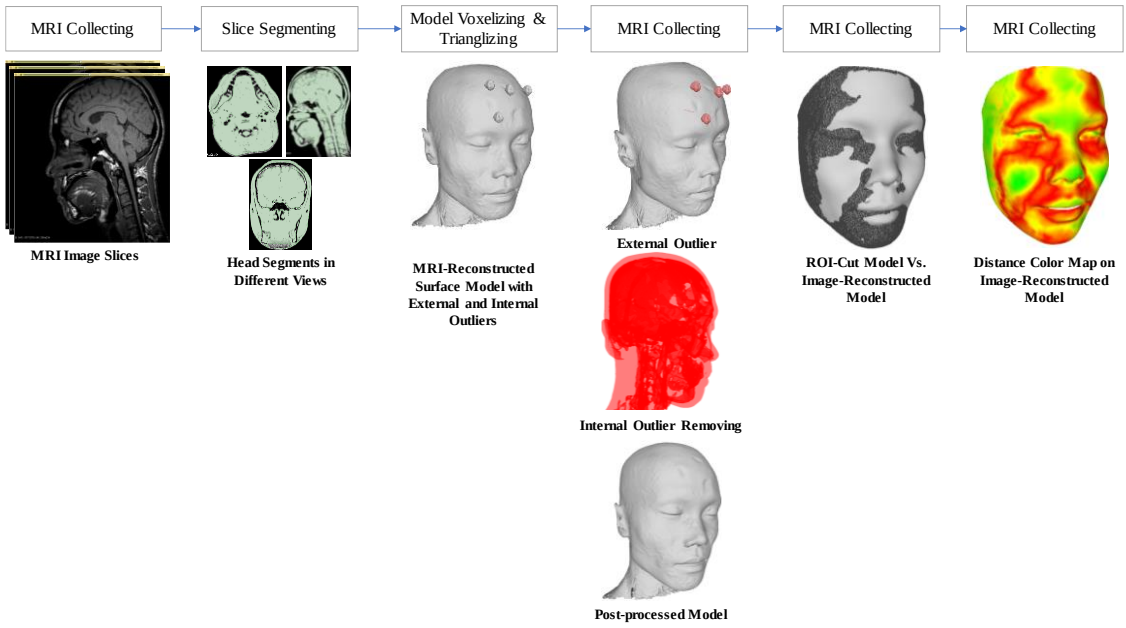


Figure 25. Reconstructed 3D face shape from the MRI images and segmentation. The 3D face shape was finally registered to the coordinate system of the image-based reconstructed face model before calculating the Hausdorff distances.

3.2. Computational results

The input images of the frontal face of the 2 facial palsy patients and 2 healthy subjects are used to reconstruct the corresponding patient-specific face. The reconstructed 3D face shapes were shown in Figure 26 with three applied methods. Comparing between three methods, the second method can reconstruct wrinkles with a full head instead of a cropped face compared with methods one and three. The second and third methods were able to reconstruct the shape detail parameters such as shape, pose, and expression, while the first method only reconstruct the subject in the neutral position.

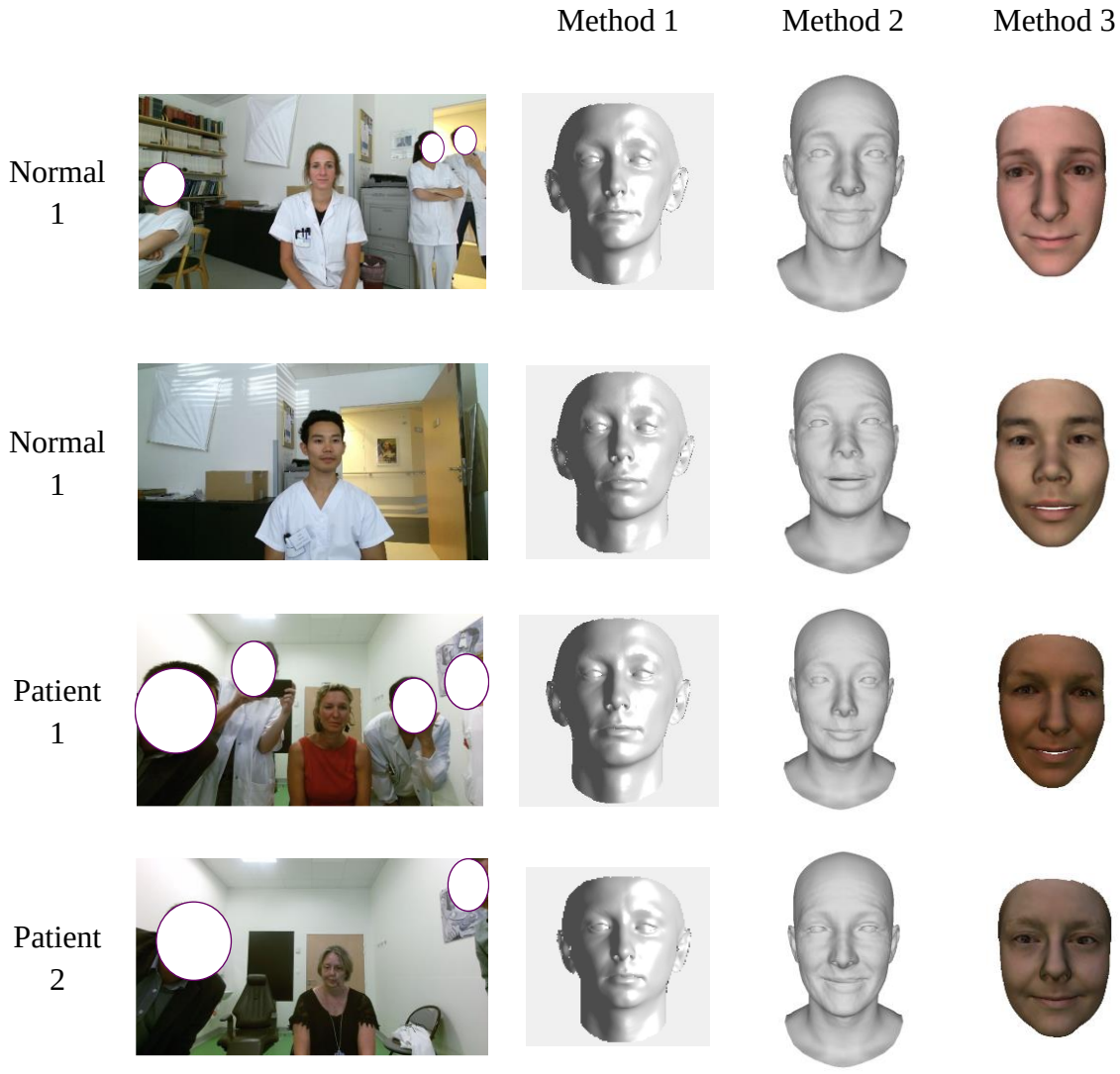


Figure 26. 3D face reconstruction from an input image

The performance then was quantified by comparing it with the 3D face obtained from the 3D camera Kinect and the MRI image. The 3D face from the MRI-based method can be treated as the ground-truth data of the person, while the 3D face from the Kinect-based method reconstructs the face with an error of about 1mm. Figure 27 demonstrates the smallest error of the 3D face reconstructed from the input image and the 3D face from the MRI-based method for the first method (fitting a 3DMM). Only three subjects (two normal

subjects and one patient) were estimated due to the MRI data of the second patient is not available. The average error of the three subjects is from 2.020 mm to 6.310 mm. The smallest error is observed in the central area of the face, while the performance suffers heavily at the jaw. This is because the input image is in the frontal of the face, while the jaw part is occluded from the frontal face image.

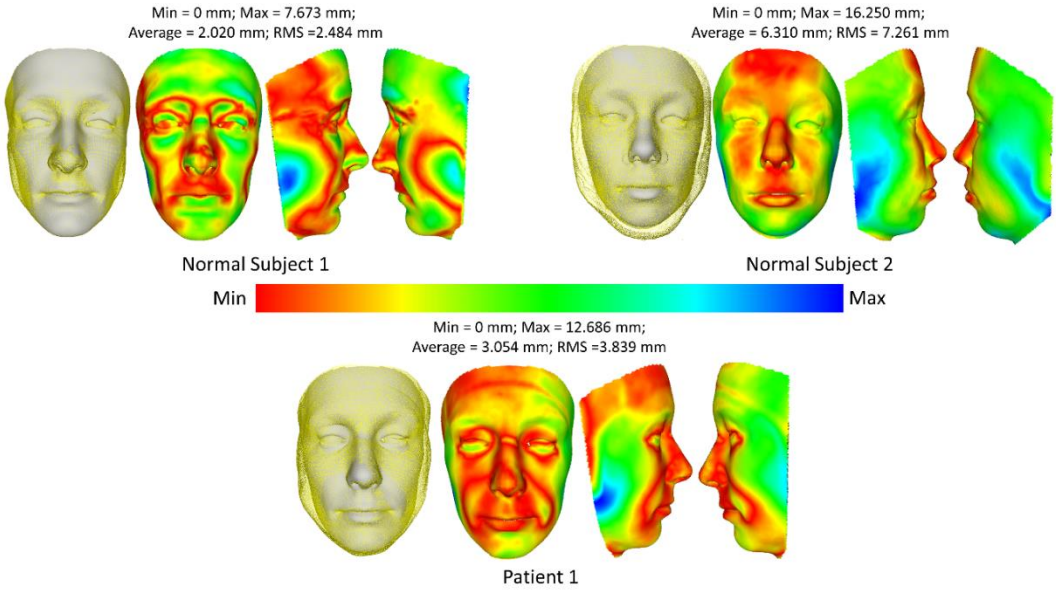


Figure 27. Comparison of 3D face reconstruction (grey) and 3D face reconstruction from MRI (yellow) using the first method (fitting a 3DMM)

The smallest errors of the 3D face reconstructed from the input image and the 3D face from the MRI-based method were illustrated in Figure 28 and Figure 29 for the second and third methods respectively. The average error of the three subjects is from 1. 7 mm to 2.5 mm. These errors for the third method range from 1.1 mm to 1.6 mm.

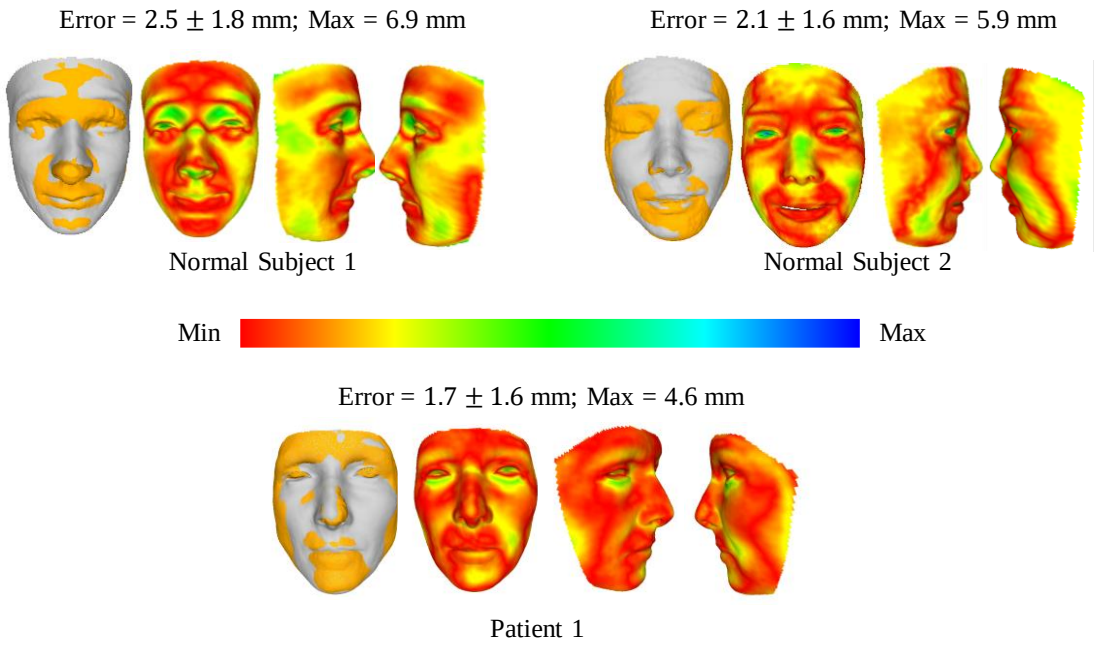


Figure 28. Comparison of 3D face reconstruction (grey) and 3D face reconstruction from MRI (yellow) using the second method (DECA)

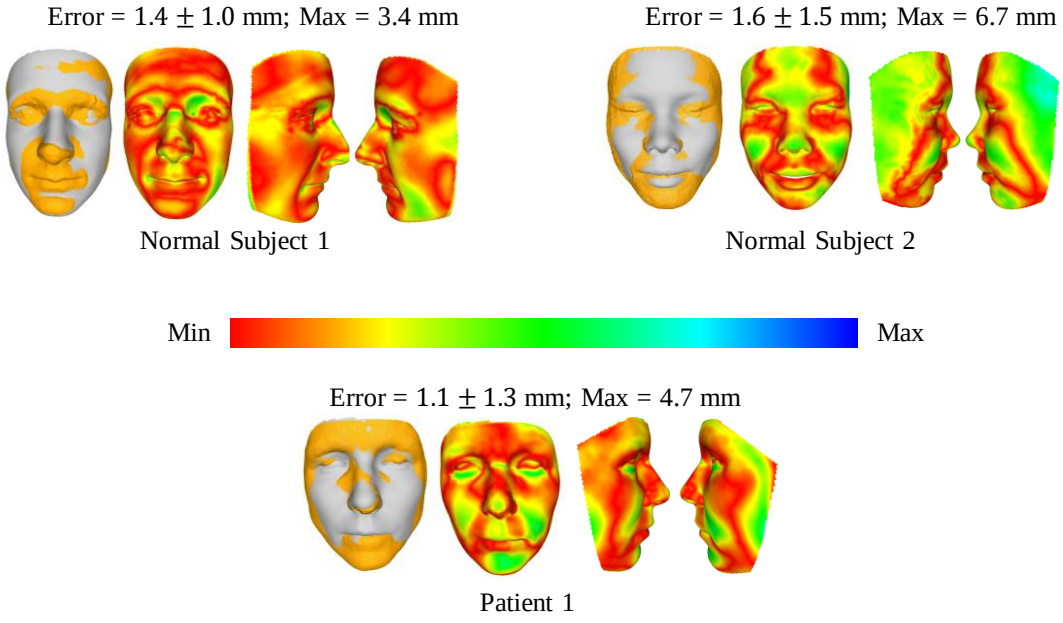


Figure 29. Comparison of 3D face reconstruction (grey) and 3D face reconstruction from MRI (yellow) using the third method (deep 3D face reconstruction)

The best prediction of the third method compared with the MRI ground truth data is 1.1 mm with a maximum error is 3.7 mm, while the worse prediction is 2.8 mm with a maximum error of 9.1mm as shown in Figure 30.

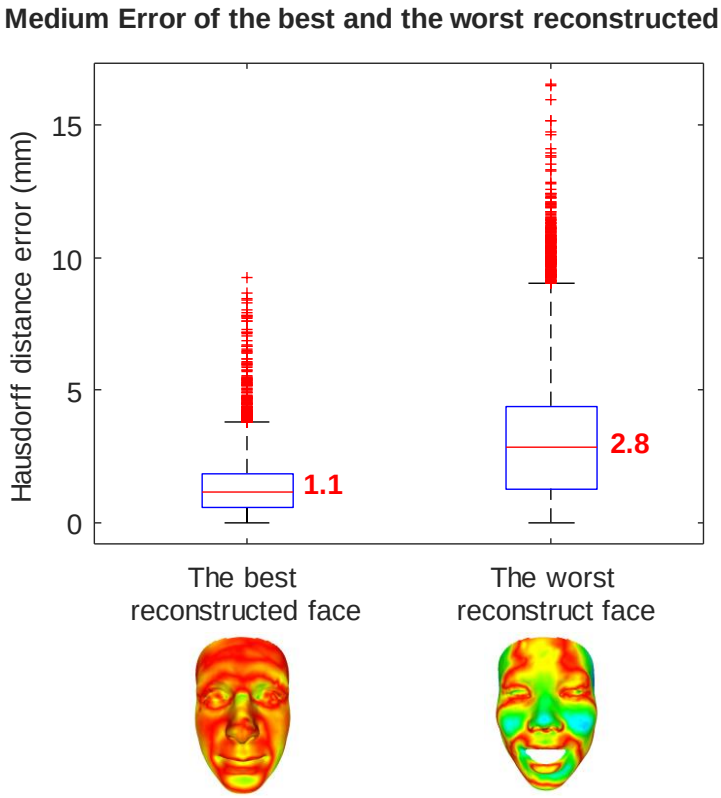


Figure 30. The error of the best and the worst prediction cases of the third method compared with MRI ground truth data

The reconstruction of a healthy subject in different mimic positions was shown in Figure 31. In the neutral position, the medium reconstruction error is 1.4 mm, while this error is 1.3 mm and 1.7 mm in smile, and [e] and [u] pronunciations respectively.

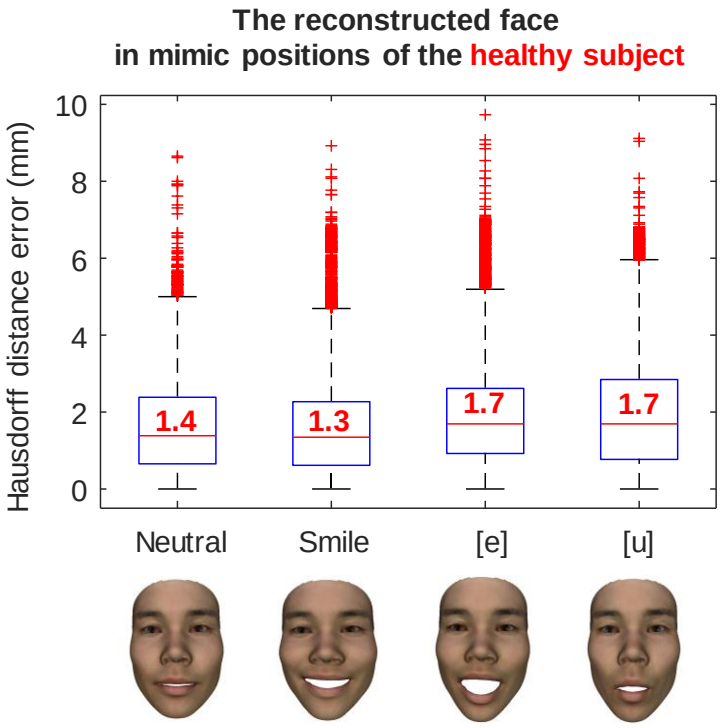


Figure 31. The error of the reconstructed face in mimic position of a healthy subject.

The reconstruction of a facial palsy patient in different mimic positions was shown in Figure 32. The medium reconstruction error is 1.1 mm, 1.4 mm, 1.3 mm, and 0.9 mm in neutral, smile, and [e] and [u] pronunciations respectively.

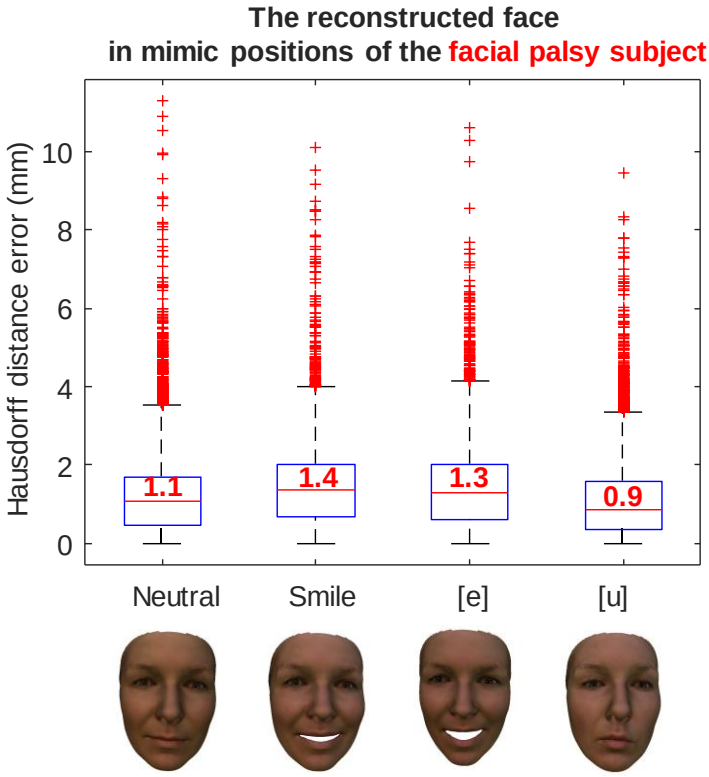


Figure 32. The error of the reconstructed face in mimic position of a facial palsy subject.

All comparison mean error ranges are reported in Table 3 for all subjects and patients in the neutral position. The mean error of all subjects from the third method is smaller than that in the second and the first methods.

Table 3. The error of the reconstruction compared with MRI and Kinect reconstructions

Method	Subject	Error (mm)	Method	Subject	Error (mm)
Fitting – Kinect comparison	1	2.3 ± 2.9	Fitting – MRI comparison	1	2.0 ± 2.5
	2	6.3 ± 7.6		2	6.3 ± 7.3
	3	2.4 ± 2.9		3	3.1 ± 3.8
	Mean	3.7 ± 4.5		**Mean**	3.8 ± 4.5
Deca – Kinect comparison	1	2.6 ± 1.9	Deca – MRI comparison	1	2.9 ± 2.1
	2	1.5 ± 1.5		2	2.6 ± 2.1
	3	1.5 ± 1.4		3	2.2 ± 2.0
	4	2.2 ± 1.7		4	1.7 ± 1.6
	Mean	1.8 ± 1.6		**Mean**	2.3 ± 1.9
Deep3Dface – Kinect comparison	1	1.7 ± 1.3	Deep3Dface – MRI comparison	1	1.6 ± 1.1
	2	1.8 ± 1.3		2	2.3 ± 1.6
	3	1.3 ± 1.0		3	1.8 ± 1.4
	4	1.4 ± 1.0		4	1.8 ± 1.5
	Mean	1.5 ± 1.1		**Mean**	1.9 ± 1.4

Several examples of 3D faces reconstructed were illustrated in Figure 32 using method 3 (deep 3D face reconstruction) for facial palsy patients from 2D images collected in open access [352].

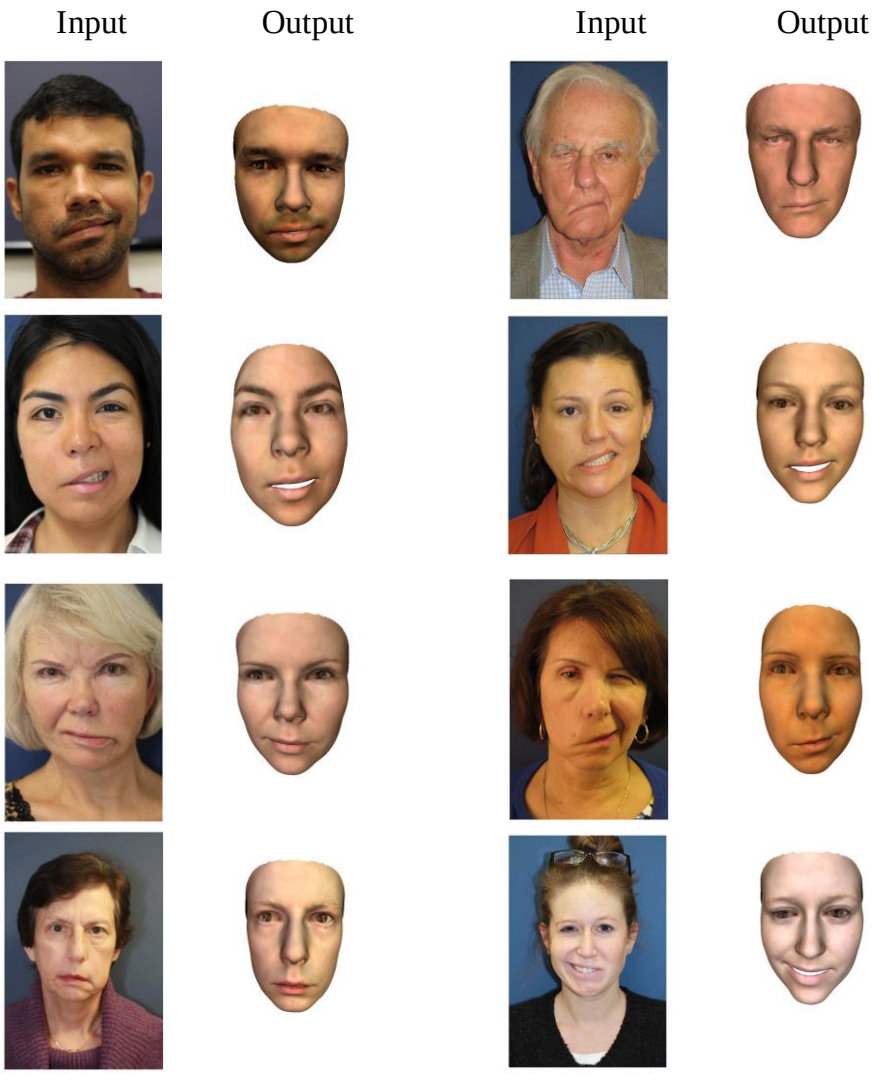
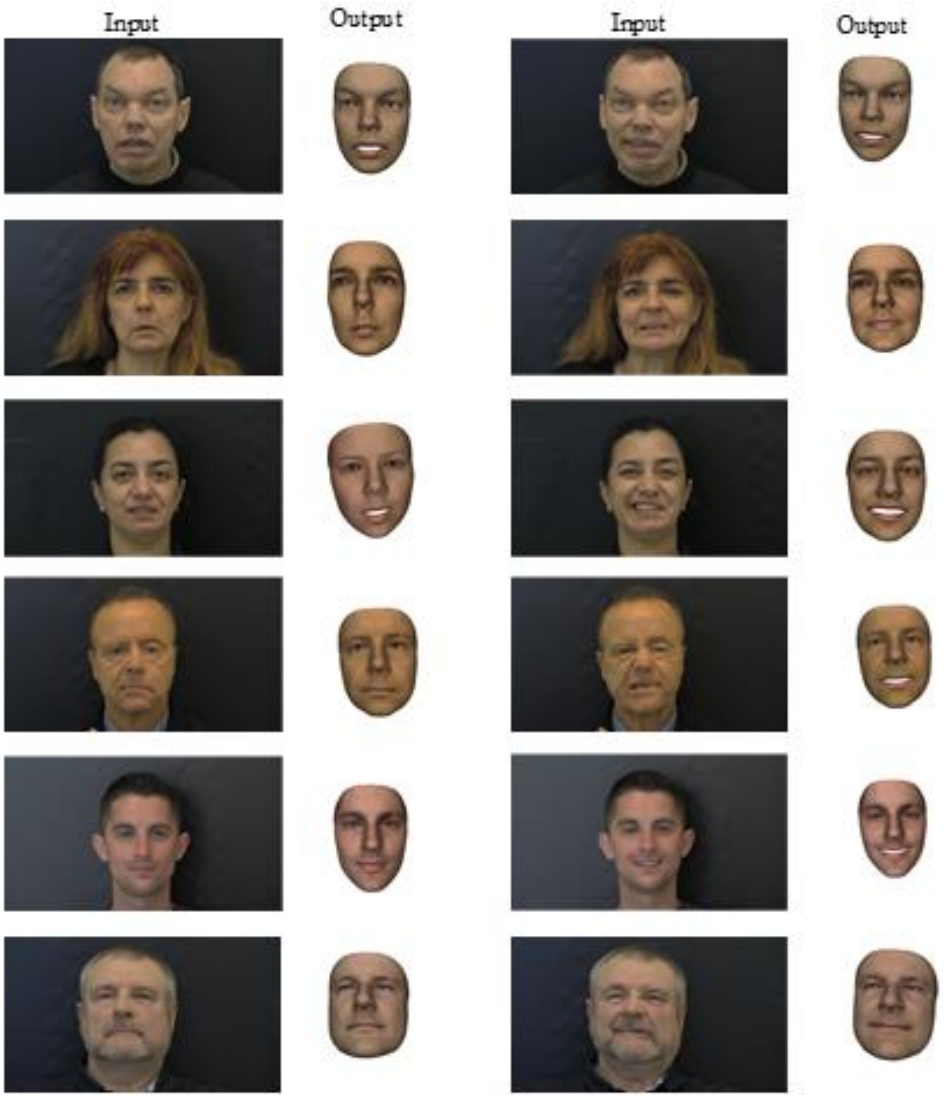


Figure 33. The 3D face reconstruction of facial palsy patients using method 3 (deep 3D face reconstruction) using collected images in open access.

The reconstructed faces of 12 patients in neutral and smiling poses from 2D images obtained at CHU Amiens were illustrated in Figure 34.



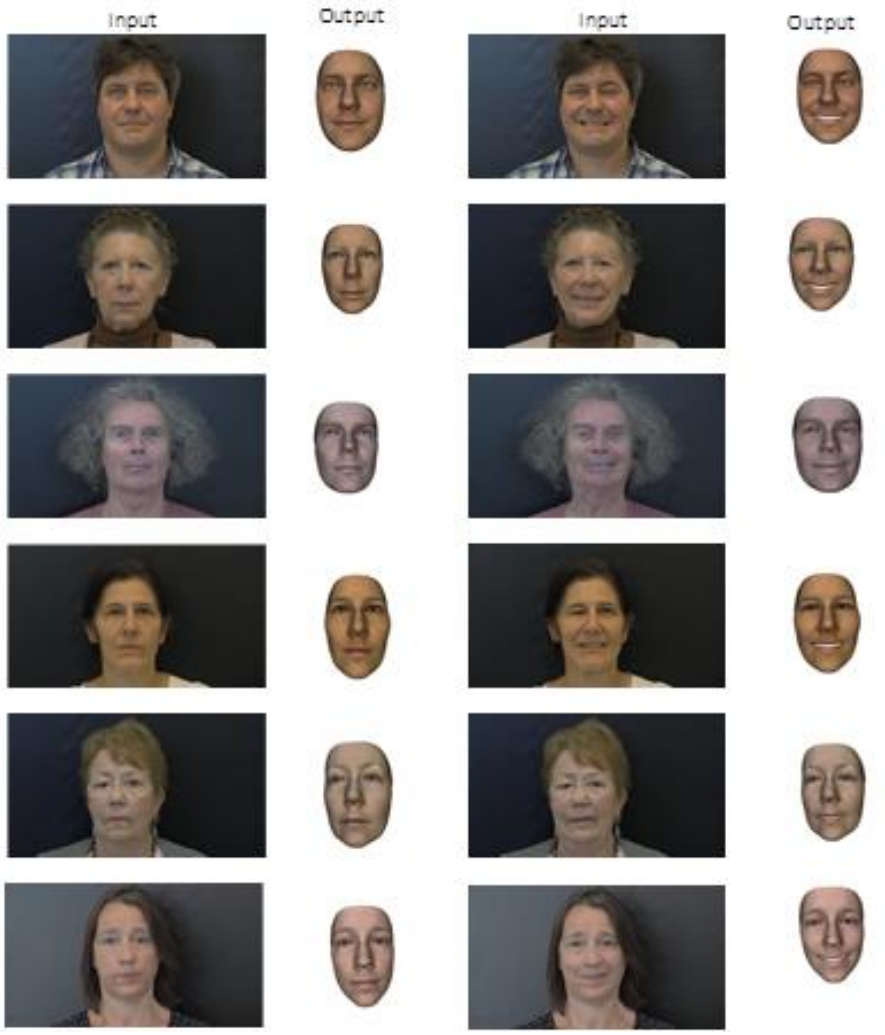


Figure 34. The 3D face reconstruction of the last 6 facial palsy patients using method 3 (deep 3D face reconstruction) using images from CHU Amiens.

For all patients with the face in neutral position (Figure 33 and Figure 34), the output reconstructed 3D face has a quite close appearance to the individual in the input 2D image. The asymmetric feature of the mouth of all patients can be observed in the reconstructed 3D faces. In the eye region, this asymmetric feature seems less noticeable. In the patients of the second dataset, the asymmetry is not much observed for both positions including neutral and smiling. This is probably due to the degree of severity of the facial palsy, which seems to be less important than the first dataset. This demonstrates the limitation of using the database with normal faces instead of facial palsy.

3.3. Discussion

Fast reconstruction of the 3D face shape plays an important role in the suitable use of computer-aided decision support systems for facial disorders. This allows to track the normal and abnormal facial deformations in static and dynamic postures leading to improve diagnosis and rehabilitation of the involved patients [33], [364]. The facial analysis for diagnosis and treatment has mainly been based on 2D images [365]–[367], which remains a challenge due to variation poses, expressions, and illumination. However, the 3D information collected from scanners and other stereo devices is time-consuming and expensive [112], [358], [368]. Recently, effective data science and deep learning methods have been developed for reconstructing 3D face information from a single image or from multiple images [128]. This opens new avenues for 3D face shape reconstruction for facial palsy patients. In the present study, we applied three state-of-the-art methods (a morphable model and two pre-trained deep learning models) to reconstruct the 3D face shape in the neutral and facial mimics postures from a single image. Obtained results showed a very good reconstruction error level by using a well pre-trained deep learning model applied for healthy subjects as well as facial palsy patients. The reconstruction is very fast and this solution is very suitable to be included in a computer-aided decision support tool.

Regarding the comparison with ground truth data from Kinect depth sensor and MRI data, the best mean errors range from 1.5 to 1.9 mm for static and facial mimic postures. These findings are in agreement with the accuracy level reported in the literature. Accuracy comparison for healthy subjects revealed that in the neutral position, the error range is less than 2mm when comparing between Basel Face Model (BFM), FaceWarehouse model, and FLAME model [358]. Moreover, the error range from 5 to 10mm for positions with large movement amplitude (mouth opening, facial expressions) [358]. In particular, all three methods estimate the 3D face model parameters without any paired ground truth data requirement. For the near frontal view image, all the methods can reconstruct the 3D face of the patient well in the central face area, however, the first method turns out badly to keep low error at the jaw where is occluded, while two other methods can handle the occlude part relatively stable. This is because the both second and third methods were trained with the loss function associated with the pose change. Interestingly, the first method reconstructs the second healthy subject, who is Asian, with a large error (~6.3 mm), while these errors are relatively smaller (2 – 3 mm) for other subjects, who are French. The reason for this is that the first method based on a 3DMM model was built from most of the subjects being Caucasian. While the second and third methods, which were based on the 3DMM model was trained from more diversity in ethnicity, there is not high different error between each subject when reconstructing the 3D face of all subjects (1.7 – 2.9 mm and 1.6-2.3 mm for the second and third methods respectively). This might prove that with more diversity in ethnicity when building the 3DMM model, the result of the reconstruction can be better. Method 3 was also applied to reconstruct 3D faces of facial palsy patients from unconstrained conditions (images were captured by any devices), since it has the lowest reconstruction error. The method is good at capturing asymmetric features in the mouth area, but less so in the eyes. This is due

to the method's usage of the FLAME model, which includes various expressions but does not include any patients with facial palsy.

In the present study, the first method fits a 3DMM to a single image based on a scale orthographic projection [129]. The method first detects facial landmarks detected in the input image, then projects the set of corresponding 3D points from the 3D model to get 2D points, and finally estimates the shape and pose parameters of the face model by minimizing the error between the 2D facial landmarks from the 3D model and 2D facial landmarks from the 2D facial image. The second used method reconstructs the 3D face based on an established FLAME head model [357]. The method is based on a resNet-50 deep learning model to learn the shape parameters such as shape, expression, pose, detail, and appearance parameters such as albedo and lighting. The model is trained by minimizing the loss function estimated from the input image and the synthesized image generated by decoding the latent code of the encoded input image. The third applied method reconstructs the 3D face of the patient with weakly-supervised learning to regress the shape and texture coefficients from a given input image [361]. This method was also based on a hybrid-level loss function to train the resNet-50 deep learning model. Regarding the 3D shape reconstruction, our findings confirmed the high accuracy level of the 3D pre-trained deep learning models for facial palsy patients. In particular, the third method was able to reconstruct the geometric details such as shape, pose, and expression while the second method reconstructs the face with the wrinkle detail. The findings revealed also that the morphable model provided a lower accuracy level. The second and third methods result in better accuracy due to integrating the loss of both geometric information (e.g. the landmark loss) and appearance information (e.g. the photometric loss), while the first method only counts the landmark loss and ignores the appearance information. Another reason for being lower accuracy is that the first method estimates the shape parameters from the Basel 3D Morphable model [133], which was only modeled by the face database of most Caucasian subjects at neutral expression, while the second and third methods were based on the 3D face models of more diversities in ethnicity and variations in facial expressions such as the FLAME model [358] and FaceWarehouse [138] respectively. Especially, the FLAME model was trained from sequences of 3D face scans that can generalize well to the novel facial data of the different subjects, which is more reliable and flexible to capture patient-specific facial shapes.

One important limitation of the present study deals with a small number of subjects and patients used for prediction. Another limitation deals with the lack of facial palsy patients in the learning database. This results in reducing several facial palsy patient's features (e.g. asymmetric face, dropping mouth corner, cheek) while reconstructing their 3D face. Thus, a larger and more diverse 3D facial database including facial palsy subjects should be acquired to confirm our findings and toward a potential clinical application. Moreover, another limitation of the study relates to the usage of the 3D statistical facial model. The first method used a 3DMM was based on the PCA basis vectors so that the reconstruction of more detailed information such as expressions and wrinkles can become a hard task. While the second and third methods improve that by building a more diverse model with subtler information such as expressions and wrinkles but still use a linear model could generate more error due to facial shape variations, which cannot be modeled perfectly using a combination of linear

components as noted in [128], [369], [370]. Improving the existing 3D face model can be a potential suggestion for future works. Another limitation relates to the effect of the variation of the 2D input image such as the pose and lighting condition have not been investigated. The variation along with the quantity of facial palsy patients are needed for improving the result of the reconstruction should be performed in the future work.

3.4. Conclusions

The 3D reconstruction of an accurate face model is essential to provide reliable feedback for clinical decision support. Medical imaging and specific depth sensors are accurate but not suitable for an easy-to-use and portable tool. The recent development in deep learning (DL) models opens new challenges for 3D shape reconstruction from a single image. However, the 3D face shape reconstruction of facial palsy patients is still a challenge and this has not been investigated.

In this present chapter, I proposed a methodology to perform 3D face reconstruction from a single 2D image captured from Kinect v2. The methodology could also be used for 2D image from any devices for reconstructing 3D face of patients with facial palsy. The methodology used several methods to reconstruct the 3D face shape models of the facial palsy patients in natural and mimic postures from one single image. Three different methods (3D Basel Morphable model and two 3D Deep Pre-trained models) were applied to the dataset of two healthy subjects and two facial palsy patients. Reconstructed outcomes showed a good accuracy level compared to the 3D shapes reconstructed using Kinect-driven reconstructed shapes ($1.5 \pm 1.1\ mm$) and MRI-based shapes ($1.9 \pm 1.4\ mm$).

The outcome of this chapter is a paper in preparation: “Fast 3D face reconstruction from a single image using different deep learning approaches for facial palsy patients” to be submitted to *Machine Vision and Applications (Q2, IF@2020=2.59).*

This present chapter opens new avenues for the fast reconstruction of the 3D face shapes of the facial palsy patients from a single image. As perspectives, reconstructed faces could be used for further analyzing the face in terms of expression and symmetry. Furthermore, the best DL method will be implemented as software as “2D/3D reconstruction” into our computer-aided decision support system for facial disorders.

Chapter 4:

Enhanced Facial Expression Recognition using 3D Point Sets and Geometric Deep Learning

Facial expression recognition plays an essential role in human conversation and human-computer interaction. Previous research studies have recognized facial expressions mainly based on 2D image processing and conventional machine learning approaches. This requires sensitive feature engineering such as key point extraction. Other studies expressed 3D face in terms of a new presentation such as curvature descriptors, SIFT feature, spatial distances, or displacements. This, however, transforms the data and often reduces the depth information on the face. The purpose of the present study was to recognize facial expressions by applying a new class of deep learning called geometric deep learning directly on 3D point cloud data.

Two databases (Bosphorus and SIAT-3DFE) were used. The Bosphorus database includes sixty-five subjects with seven basic expressions (i.e. anger, disgust, fear, happy, sad, surprise, and neutral). The SIAT-3DFE database has 150 subjects and 4 basic facial expressions (neutral, happiness, sadness, and surprise). Firstly, pre-processing procedures such as face center cropping, data augmentation, and point cloud denoising were applied to 3D face scans. Secondly, a geometric deep learning model called PointNet++ was applied. A hyperparameter tuning process was performed to find the optimal model parameters. Finally, the developed model was evaluated using the recognition rate and confusion matrix.

This part corresponds to task number 2 in the workflow (Figure 35).

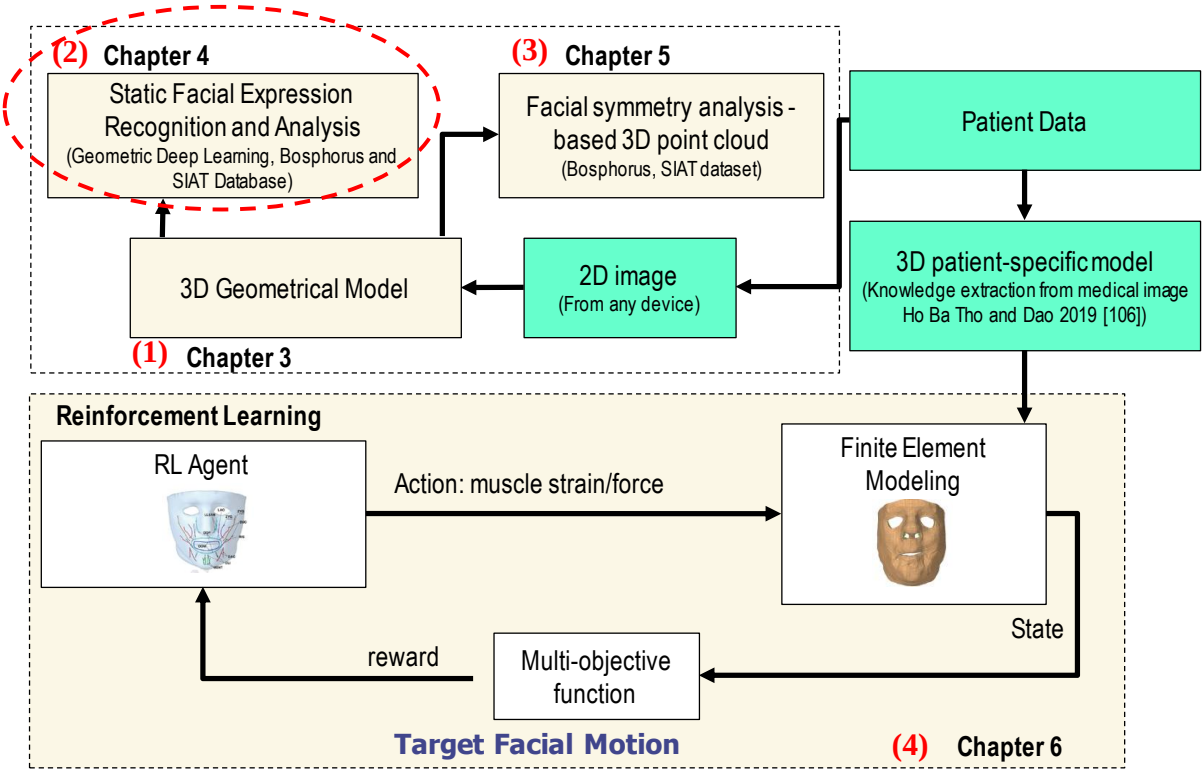


Figure 35. The thesis framework: the part corresponds with facial expression recognition

4.1. Materials and methods

4.1.1. Learning Databases

The Bosphorus database [113] was used for facial expression recognition. This database was acquired by using Inspeck Mega Capturor II 3D device. The database consists of 105 subjects (61 males and 44 females) in various poses, expressions, and occlusion conditions. Most of the subjects are Caucasian and their ages range from 25 to 35 years old. Each subject was scanned from 31 to 54 times with different expressions and action units (AUs). In total, the database has 4666 face scans. For each face scan, 24 facial landmarks have been manually labeled as shown in Figure 36. Seven basic emotional expressions such as happiness, surprise, fear, sadness, anger, disgust, and neutral will be processed for facial recognition.

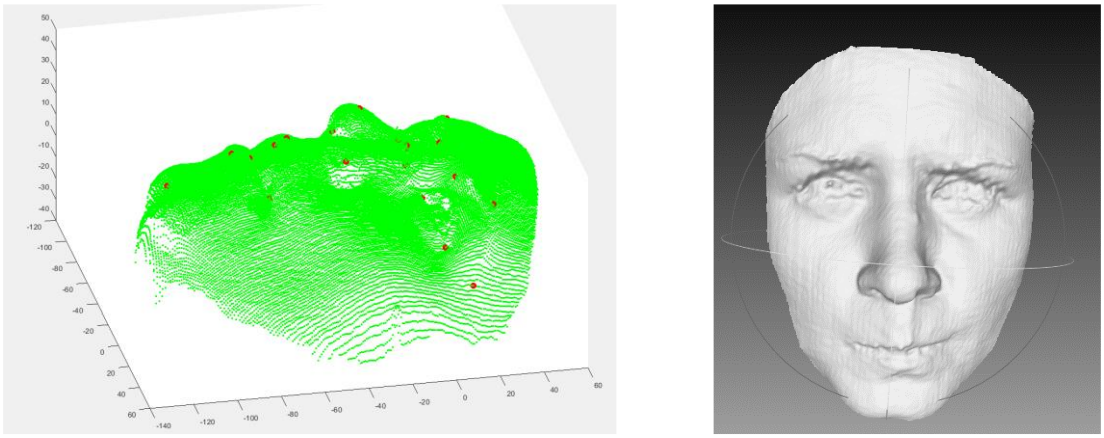


Figure 36. 3D face with facial landmarks: 3D facial point cloud with 24 facial landmarks (left) and associated mesh of 3D face scan (right).

The SIAT-3DFE database [201] was also applied for facial expression recognition. This high-resolution database was acquired by two structured light scanner systems called CASZM-MVS600. The data consists of 8000 3D facial scan models from 500 Asian volunteers aged from 18 to 50. Each participant was scanned 16 times with 4 basic facial expressions such as neutral, happiness, sadness, and surprise.

4.1.2. Data pre-processing procedures

From the acquired Bosphorus database, 65 subjects were selected for recognizing seven basic expressions such as anger, disgust, fear, happy, sad, surprise, and neutral. Note that other subjects were excluded due to a lack of some facial expressions. Totally, 455 original face scans were used for recognizing facial expressions. The raw face scans were preprocessed using statistical techniques [371] to remove outliers and reduce noise. The face scans were then cropped in the center region using facial landmarks such as outer eyebrow, nose tip, and lower lip outer middle. The data augmentation strategy was done by randomly sampling points of 3D face point cloud. After data augmentation, the data size increases to

14560 facial point clouds. The point cloud data was also translated with the center of the face at the origin of coordinate and normalized to the range from -1 to 1.

Similarly, 150 subjects from the SIAT-3DFE database including 75 males and 75 females were selected with four basic expressions such as neutral, happiness, sadness, and surprise. Other subjects were excluded due to missing data and privacy disagreements. The used pre-processing produces were the same as those used for the Bosphorus database. Thus, after data augmentation, the data size reaches 9600 facial point clouds from 600 original facial scans.

4.1.3. Geometric deep learning model

The recognition of facial expressions was performed by implementing a hierarchical neural network called PointNet++. PointNet++ builds a hierarchical grouping of points and abstract larger local regions along the hierarchy. The main architecture of PointNet++ comprises of three key layers: 1) *Sampling layer* selects a set of points from input points, which defines the centroids of local regions; 2) *Grouping layer* constructs local regions sets by finding neighboring points around centroids; and 3) *PointNet layer* to encode local regions into feature vectors. These feature vectors are then used to perform the classification or segmentation task.

PointNet used symmetry functions takes *n* vectors as input to aggregate the information and output a new vector that is invariant to the input order (Figure 37). From an unordered point set $\{x_1, x_2, \dots, x_n\}$ with $x_i \in \mathbb{R}^d$, a set function f that transforms a set into a feature vector.

$$f(x_1, x_2, \dots, x_n) = \gamma\left(\max_{i=1,\dots,n}\{h(x_i)\}\right)$$

where γ and h are multi-layer perceptron networks (MLP). PointNet learns features independently by applying MLP layer for each point and extracts global features with a max-pooling layer.

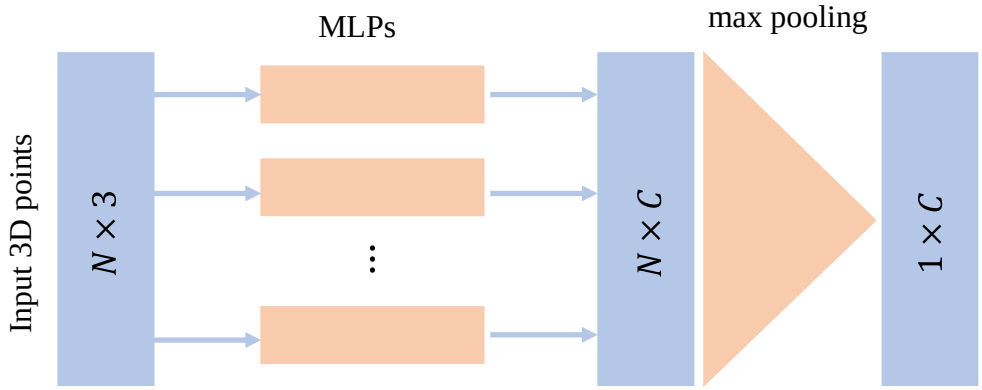


Figure 37. PointNet feature extraction: N is the input number of 3D points, C is the number of learning features, and MLPs is the Multi-Layer Perceptron.

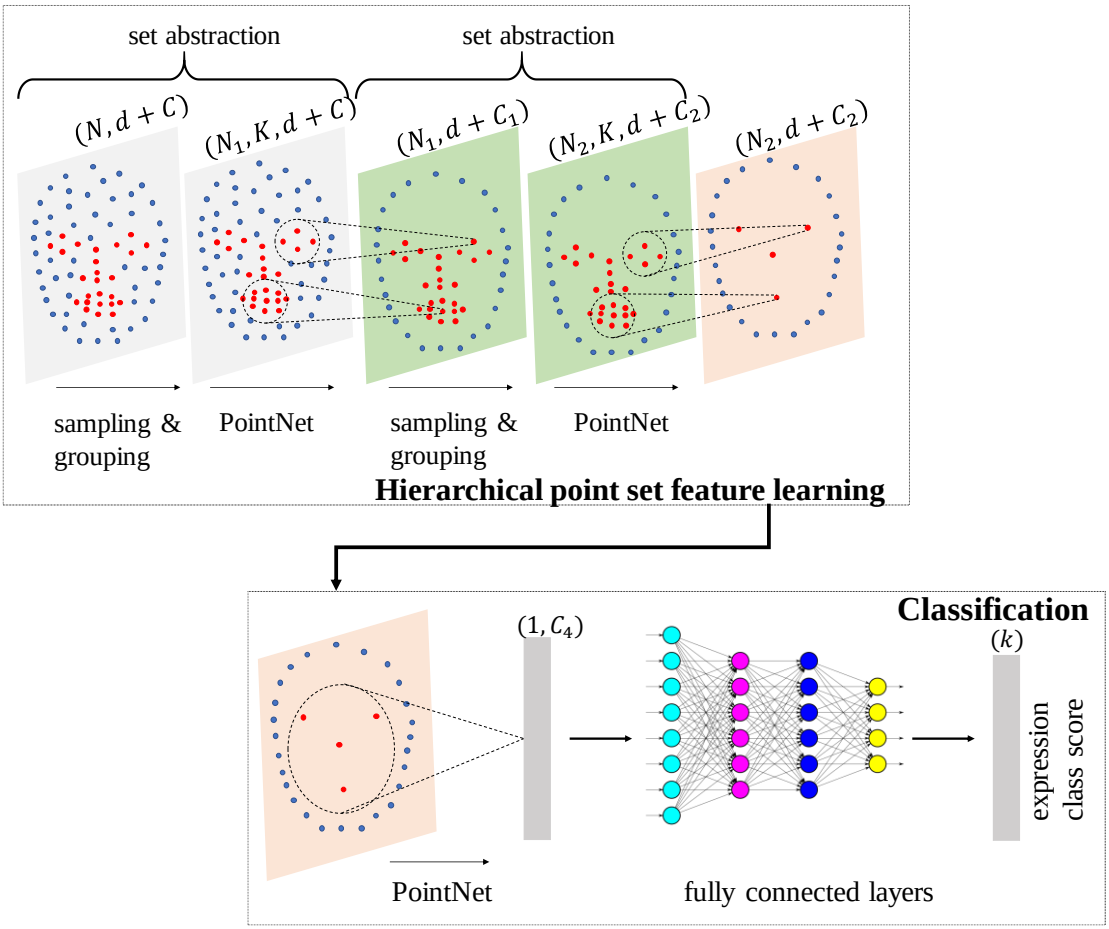


Figure 38. The hierarchical local feature learning architecture for the classification task was illustrated in 2D space as an example. N, N_1, N_2 are number of points in d dimensional coordinates and C, C_1, C_2 dimensional point feature for each processing step. K is the number of points within a radius from centroid points. k is the number of expressions that need to be recognized.

Due to the independent learning feature for each point, PointNet cannot capture the local structural information between points in local regions. PointNet++ copes with this issue by adding a sampling layer and grouping layer to generate local regions from the neighborhood of each point. The local feature is then learned from local regions by PointNet layer.

The network is built as following architecture:

$$SA(512, 0.2, [64, 64, 128]) \rightarrow SA(128, 0.4, [128, 128, 256]) \rightarrow SA([256, 512, 1024]) \rightarrow FC(512, 0.5) \rightarrow FC(256, 0.5) \rightarrow FC(k)$$

where the network comprises two set abstraction levels $SA(K, r, [l_1, \dots, l_d])$ with K local regions of search radius r. Each set abstraction level uses PointNet with width $l_i (i = 1, \dots, d)$ of d fully connected layers. A global set abstraction level, named $SA([l_1, \dots, l_d])$ coverts set to a single vector. Two fully connected layers $FC(l, d_p)$ is with width l and dropout regularization ratio d_p are also used. A fully connected layer $FC(k)$ is applied for classifying k classes of facial expression.

PointNet++ model was trained and tested on a free for public use internal research tool called Collaboratory. It is a jupyter notebook environment with a 12GB NVIDIA Tesla K80 GPU, CUDA version 10.1, the Tensorflow version 1.15.2, and Python version 3.6.

4.1.4. Accuracy evaluation

Regarding the evaluation of the proposed model, five-fold cross-validation was applied. We divided the learning dataset into 70% for training, 10% for validating, and 20% for testing. A hyperparameter tuning process was performed to select the optimal values of the PointNet++ model parameters. Precisely, our hyperparameter tuning process works by manually adjusting each parameter, while remaining all other parameters unchanged. Multiple trials were performed in a single training task for facial expression recognition. Each trial was completed with values within a set of chosen hyperparameters. The recognition rate was then estimated and compared to find the effective hyperparameter values. Related ranges of values of these model parameters were selected as follows: batch size (6, 16, and 32), number of epochs (41, 81, and 101), learning rate (0.0001, 0.001, and 0.01), decay step (100000, 200000, and 300000), decay rate (0.35, 0.7, and 1.4), optimization method ('Adam' and 'Momentum'), and point cloud density (128, 256, 512, 1024, 2048, and 4096).

The training process was monitored using the Tensorboard module. The confusion matrix was also evaluated. The accuracy equals the total correct predictions divided by total predictions. The accuracy of the prediction was the mean accuracy of five prediction times. The mathematical equation of this metric is expressed as follows:

$$Acc = \frac{Total\ correct\ predictions}{Total\ predictions}$$

4.2. Computational results

4.2.1. Hyperparameter tuning process

The accuracy rate of facial expression recognition of hyperparameter tuning on the Bosphorus database is shown in *Table 4*. Figure 39 shows the effect of point cloud resolution on the recognition rate. It is important to note that the higher point density results in higher recognition accuracy. The model used a smaller batch size (8), larger learning rate (0.01) at a decay rate of 0.07, and decay step 200000, trained with 81 epochs, and used Adam optimization showed a better accuracy level. The accuracy is higher for smaller batch sizes. This behavior is reasonable because a larger batch size degrades the quality of the model due to convergence to sharp the minimizers of the training function [372]. In addition, the computational time of the method applied to the Bosphorus database is around 2 hours and 31 minutes for training after 101 epochs and 18 seconds for testing of 1-fold of the data. Regarding the SIAT-3DFE database, the computational time is 1 hour and 21 minutes for training after 101 epochs and 4 seconds for training and testing 1-fold respectively.

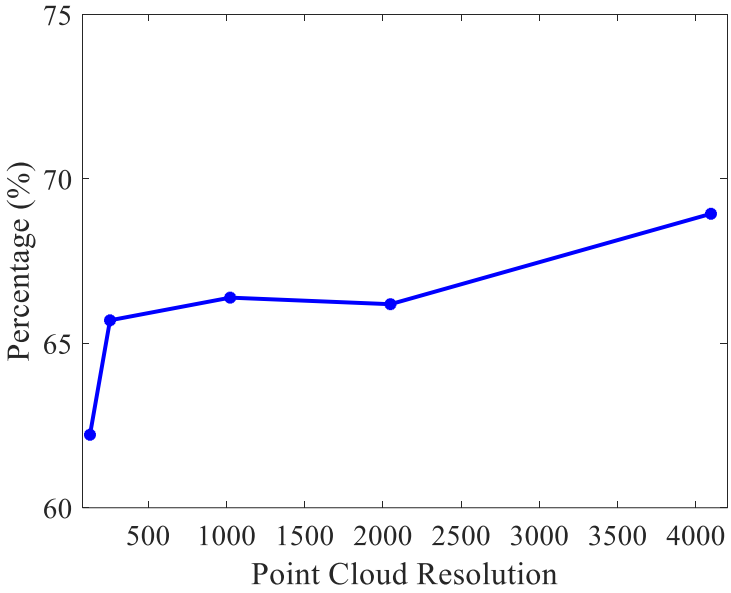


Figure 39. Facial expression recognition accuracy according to the point cloud resolution.

Table 4. The recognition accuracy of hyperparameter tuning process

Hyperparameters		Accuracy
Batch size	8	**69.01%**
	16	68.22%
	32	66.02%
Number of epochs	41	67.47%
	81	**69.01%**
	101	68.51%
Learning rate	0.0001	62.33
	0.001	68.22%
	0.01	**68.68%**
Decay rate	0.035	66.05%
	0.07	**68.22%**
	0.14	62.30%
Decay step	100000	66.15%
	200000	**68.22%**
	300000	67.18%
Optimizer	Adam	**68.22%**
	Momentum	66.51%

4.2.2. Model evaluation

The loss and accuracy metrics computed during training and validation processes with the Bosphorus and SIAT-3DFE databases were shown in Figure 40 and Figure 41 respectively. The loss rapidly drops below 0.25 and 0.15 for the training process with Bosphorus and SIAT-3DFE databases respectively. Then it keeps slightly decreasing when the training process continues. The accuracy of each case keeps increasing. The convergence of accuracy and loss metrics can be observed after 60 epochs for training with the Bosphorus database and after 80 epochs for training with the SIAT-3DFE database. For both databases, the accuracy of the validation process is lower than that of the training process, while the loss metric in validation is higher than that in training. The discriminant value between testing data and training data is higher in the SIAT-3DFE database than that in the Bosphorus database.

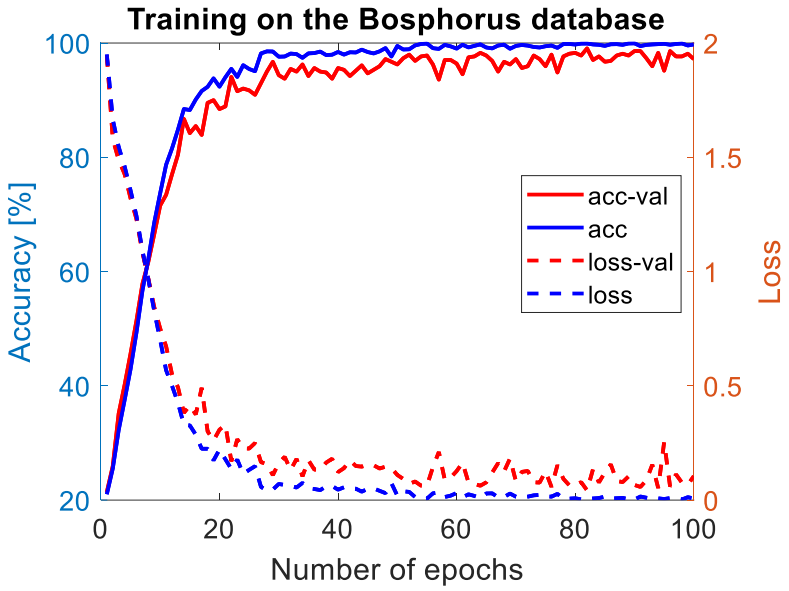


Figure 40. Accuracy and loss during training the Bophorus database with 4096 points

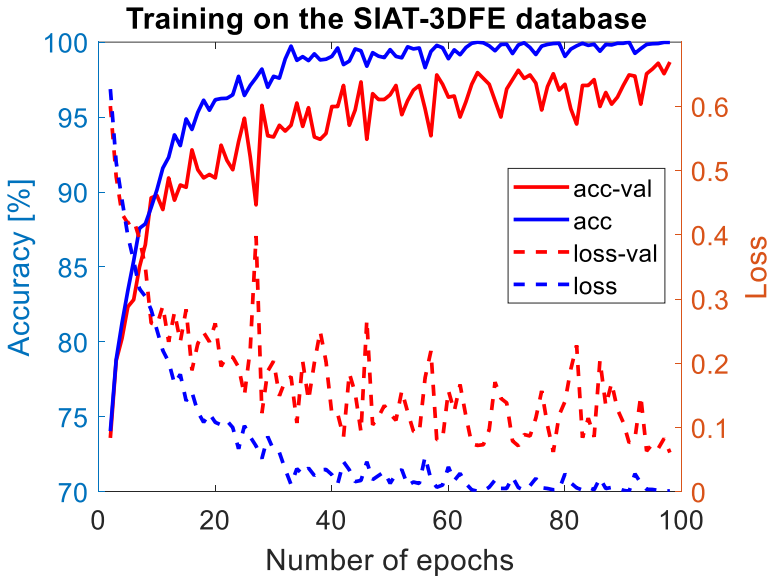


Figure 41. Accuracy and loss during training the SIAT-3DFE database 4096 points

The accuracy is 69.01% for 7 expressions and reaches 85.85% when recognizing 5 expressions like anger, disgust, happiness, surprise, and neutral. The confusion matrix of recognizing facial expressions for all 7 expressions was shown in Figure 42. Fear and sadness are the two most difficult expressions to recognize with only 57.3% and 48.6% of accuracy. From the figure, 16.1% of anger expression was recognized as sad, 19.1% of fear expression was recognized as surprise, 21.4% of sad expression was recognized as anger, and 19.7% of surprise expression was recognized as fear expression.

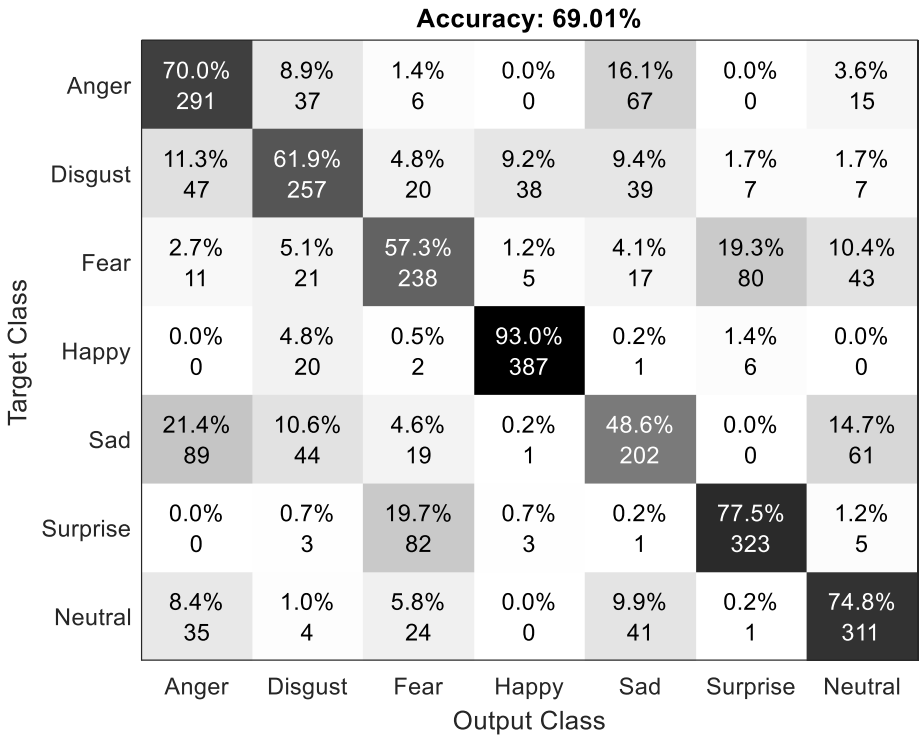


Figure 42. Confusion matrix of facial expression recognition with the Bosphorus database for 7 expressions.

The average rate of expression recognition increases significantly to 85.85% when recognizing only 5 expressions as illustrated in Figure 43. Two expressions, fear and sad were removed.

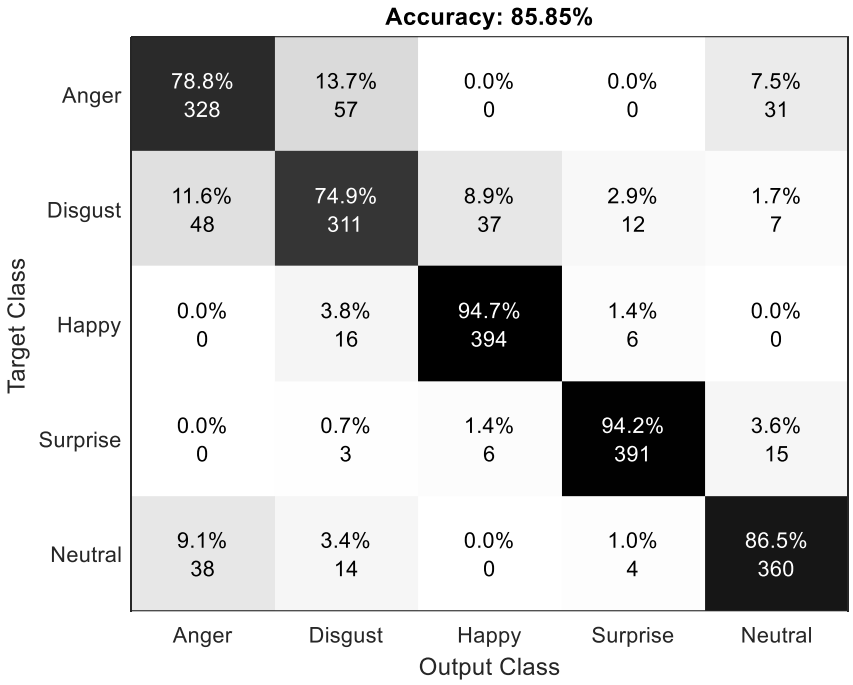


Figure 43. Confusion matrix of facial expression recognition with the Bosphorus database for 5 expressions (anger, disgust, happiness, surprise, and neutral).

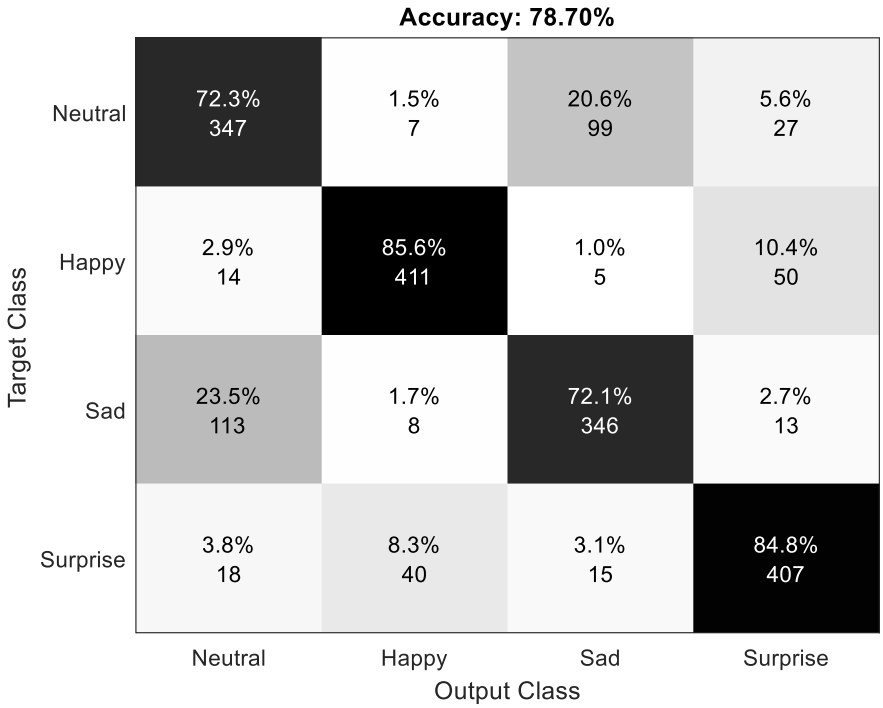


Figure 44. Confusion matrix of facial expression recognition with the SIAT-3DFE database for 4 expressions (neutral, happiness, sadness, and surprise).

The average rate of expression recognition of the SIAT-3DFE database was shown in Figure 44. This metric reaches a value of 78.70% when recognizing 4 expressions such as neutral, happiness, sadness, and surprise. From the figure, 20.6% of neutral expression was recognized as sad, 23.5% of sad expression was recognized as neutral, and 10.4% of happiness was recognized as surprise.

The comparison between our findings with the other state-of-the-art studies on the same Bosphorus database is shown in *Table 5*. Our method shows a better result than the previous study using Zernike moment feature [373]. However, the present study with 7 expressions showed a lower recognition rate than other studies using enhanced learning features such as 3D curvature descriptors [271], curvature-based descriptors [374], or covariance matrices of descriptors [375].

Table 5. Comparison with other studies on the Bosphorus database

Studies	Classifiers	Recognition rate	Salient features and number of recognized expressions
Vretos et al. [373]	SVM	60.5%	Zernike moment feature, 6 expressions
Azazi et al. [271]	SVM	79%	3D curvature descriptors, 7 expressions
Wang et al. [374]	SVM	76.56%	Curvature-based descriptors, 6 expressions
Hariri et al. [375]	SVM	86.17%	Covariance matrices of descriptors, 7 expressions
This present chapter	PointNet++	69.01%	3D point cloud, 7 expressions

4.3. Discussion

Facial expression recognition is a complex engineering task for computer vision systems. Which expressions are easy to be recognized and which are still challenging remain open questions in the computer vision community. Previous studies tried to respond to these questions by using different approaches and applications such as assisting mental or physical disease diagnosis and assessment [376], and improving the smart healthcare system [377]. Conventional machine learning and deep learning models have been commonly applied [7]. However, a fully automatic recognition system with a good accuracy level is still challenging. The present study proposed automatically facial expression recognition based on a geometric deep learning model, called PointNet++. Geometric deep learning refers to a term of deep learning that generalizes neural network models to interpret non-Euclidean data such as graphs and manifolds, rather than traditional deep learning dealing with Euclidean data such as images, text, or audio [378]. Geometric deep learning allows exploring more complex types of data as non-Euclidean data including investigating the edge of the pixels, graphs, and 3D objects, which offers the system to learn remarkable insight information about the relationship among and between pixels [379]. For recognizing facial expressions, geometric deep learning allows learning features directly from the 3D face scan. Our study is based on the processing of the 3D face scan to overcome the challenging problem of extreme pose variations and 2D image illumination variations. The result on the Bosphorus database shows that happiness, surprise, and neutral are the most recognizable expressions. Fear and sadness are the most two challenging expressions to recognize. This is compatible with subjective recognition when people often confuse between fear and surprise, and objective recognition when classifiers are often confused between sadness and anger [380], [381]. Moreover, fear and surprise share several same action units with Inner Brow Raiser, Outer Brow Raiser, Upper Lid Raiser, and Jaw Drop actions [382]. Precisely, fear expression is defined by the following AUs (Inner Brow Raiser, Outer Brow Raiser, Brow Lower, Upper Lid Raiser, Lid Tightener, Lip Stretcher, Jaw Drop), and surprise expression deals with Inner Brow Raiser, Outer Brow Raiser, Upper Lid Raiser, and Jaw Drop). Another reason for failing prediction is the shape variation. There are 12/65 subjects that produce less than 55% recognition rate. 5/12 of those subjects having the bear, which significantly affects the shape of the face results in producing a poor recognition rate. Regarding the use of the SIAT-3DFE database, the

recognition rate for 4 expressions does not outperform that of the use of the Bosphorus database for 7 expressions and 5 expressions. It is important to note that the Bosphorus database contains 27 professional actors/actresses resulting in a better distinction between expressions. Thus, the quality of the experimental data leads to a higher recognition rate. Furthermore, it is important to note that our study is the first to provide an accuracy level for facial expression recognition with the SIAT-3DFE database. Thus, our findings could be used in a future benchmark study with this database.

In addition, the recent development of novel sensing devices (i.e. Kinect, 3D scanner, Lidar) results in the availability of non-Euclidean data such as graphs, meshes, and 3D point clouds. This triggers the development of geometric deep learning leading to considerable achievements in robotics, autonomous driving, and augmented reality, and potentially playing a vital role in supporting clinicians on diagnosis and prediction. For example, three-dimensional convolutional neural networks (3D CNNs) were developed [383] for 3D medical imaging recognition, classification, detection, and segmentation. However, deep learning on 3D point clouds comes with several challenges because point cloud is irregular sparse and dense regions, unstructured and unordered [301]. The state-of-the-art PointNet++ overcomes these challenges by applying deep learning directly to the 3D point cloud [303]. In this present study, the PointNet++ was selected for our facial expression recognition and obtained results confirmed the robustness and the accuracy of this approach for processing directly complex 3D geometries. Precisely, PointNet++ model worked directly on 3D raw face point clouds to recognize seven basic facial expressions. The recognition rate was 69.01% when recognizing 7 expressions, which is better than even subjective evaluation with 64.87% [381]. The result also outperformed the previous research that was processed on the same database with 60.5% [373]. However, the present study showed a lower accuracy than several previous methodologies applied to the Bosphorus database and based on 3D curvature descriptor [271], primary facial landmarks by projecting 3D data into the 2D plane [374], and feature points extraction for finding covariance matrices of descriptors [34] with 79%, 76.56%, and 86.17% respectively. The superior level of accuracy of these studies can be explained by the fact that enhanced learning features were processed, extracted, and used for facial expression recognition. For example, Azazi et al. [271] (2015) applied the conformal mapping technique to map 3D texture face images into the 2D plane and selected the optimal facial features based on detecting seven facial landmarks at eye corners, mouth corners, and nose tip. Thus, the proposed method is still required to extract feature facial landmarks. Wang et al. [374] (2013) encoded curvature information by applying local curvature patterns and descriptors such as principal curvature, mean curvature, and shape index. Even achieved a better recognition rate than ours, this method requires handcraft feature extraction for estimating the curvature-based descriptors. Furthermore, Hariri et al. [375] (2017) used covariance matrices of descriptors to reach a very good accuracy level for facial expression recognition. However, this approach is still based on feature extraction instead of directly handling 3D point cloud data. Thus, our present study shows a comparable level of accuracy with these state-of-the-art methods without feature engineering processes. Especially, the accuracy level reaches a value of 85.85% when recognizing 5 expressions like anger, disgust, happiness, surprise, and neutral. This opens new avenues for recognizing the facial

expressions directly with 3D point clouds from novel sensing devices (e.g. Kinect V2, Azure Kinect).

Despite PointNet++ can work directly on the 3D point cloud, 3D raw point clouds still require several preprocessing tasks such as removing the noise and cropping the face in the center region and data normalized. Cropping point cloud into the center region of the face reinforces the discriminative capability of expression feature for the deep learning model. In addition, the augmentation technique is also applied to enhance the size of our database. Moreover, a hyperparameter tuning process is monitored for selecting the best parameter for learning the model of facial expression recognition. All these techniques belong to the best practices for developing and evaluating a robust deep learning model.

One of the possible ways to improve the generalization performance of the deep learning model is to use transfer learning [384]. However, in the present study, we performed cross-database learning and testing and obtained results that showed a lower recognition rate. More precisely, we trained the model on the Bosphorus database and tested it on the SIAT-3DFE database, and inverted this process. The recognition rate of cross-database testing from trained the Bosphorus database is only 24.74%. This rate reaches a value of 27.32% for training on the SIAT-3DFE database. It is interesting to note this behavior. In fact, the used PointNet++ model is database-dependent for the facial expression recognition problem. In particular, the quality of the experimental data will lead to a higher accuracy level.

The main limitation of the present study relates to the size of the used the Bosphorus database with high-quality data. Only 65 subjects with a total of 455 face scans were used for facial expression recognition, which can narrow the generalization of the method. **Therefore, to be potentially applied for clinical applications, a larger 3D facial database including the facial palsy patient's database is required for confirming and improving the obtained result.** In particular, a robust data acquisition protocol should be established to provide high-quality data for learning and testing. Furthermore, scanning 3D faces is a complex, time-consuming process. As a result, fully automatic methodologies from 2D image to 3D point cloud should be investigated to be able to apply to a rehabilitation program for facial palsy patients.

4.4. Conclusions

Facial expression recognition plays an essential role in human conversation and human-computer interaction as well as to assist doctors in diagnosing facial palsy patients. Previous research studies have recognized facial expressions mainly based on 2D image processing requiring sensitive feature engineering and conventional machine learning approaches. The purpose of the present chapter was to recognize facial expressions by applying a new class of deep learning called geometric deep learning directly on 3D point cloud data.

Two databases (Bosphorus and SIAT-3DFE) were used. The Bosphorus database includes sixty-five subjects with seven basic expressions (i.e. anger, disgust, fearness, happiness, sadness, surprise, and neutral). The SIAT-3DFE database has 150 subjects and 4 basic facial expressions (neutral, happiness, sadness, and surprise). Firstly, pre-processing

procedures such as face center cropping, data augmentation, point cloud denoising were applied to 3D face scans. Secondly, a geometric deep learning model called PointNet++ was applied. A hyperparameter tuning process was performed to find the optimal model parameters. Finally, the developed model was evaluated using the recognition rate and confusion matrix. Obtained results showed a better accuracy level according to the state-of-the-art methods using conventional machine learning.

The outcome of this chapter has been published: "*Enhanced Facial Expression Recognition using 3D Point Sets and Geometric Deep Learning*" in *Medical & Biological Engineering & Computing (Q2, IF@2020 = 3.05)* [127] (https://doi.org/10.1007/s11517-021-02383-1).

In perspective, using the database of facial palsy patients, we can quantify different expressions for facial mimics such as for example smile or pronunciation of "o", "pou" … [338] to diagnose the degree of the severity of facial palsy in collaboration with clinicians. The developed model will be integrated into REHAB_DEEPFACE as a part of the diagnosis of the severity of facial palsy.

Chapter 5:

Facial Symmetry Analysis based on Novel Shape Descriptors between Two Different Populations

Facial symmetry indicates the balance and equality of facial structure in terms of shape, size, location, and arrangement of left and right components on the sagittal plane. In fact, facial symmetry analysis is important to support the correct assessment and diagnosis of facial palsy and facial transplantation patients. This could help the doctor to design efficient individual rehabilitation programs. Previous studies have mainly been based on either subjective methods conducted by doctors or automatic systems but require facial landmark detection. Thus, the present chapter extracts a novel point descriptor from 3D point cloud data that supports analyzing the face in terms of facial symmetry. Two databases were used corresponding with two different populations (Caucasian and Asian).

This part corresponds with task number 3 (Figure 45).

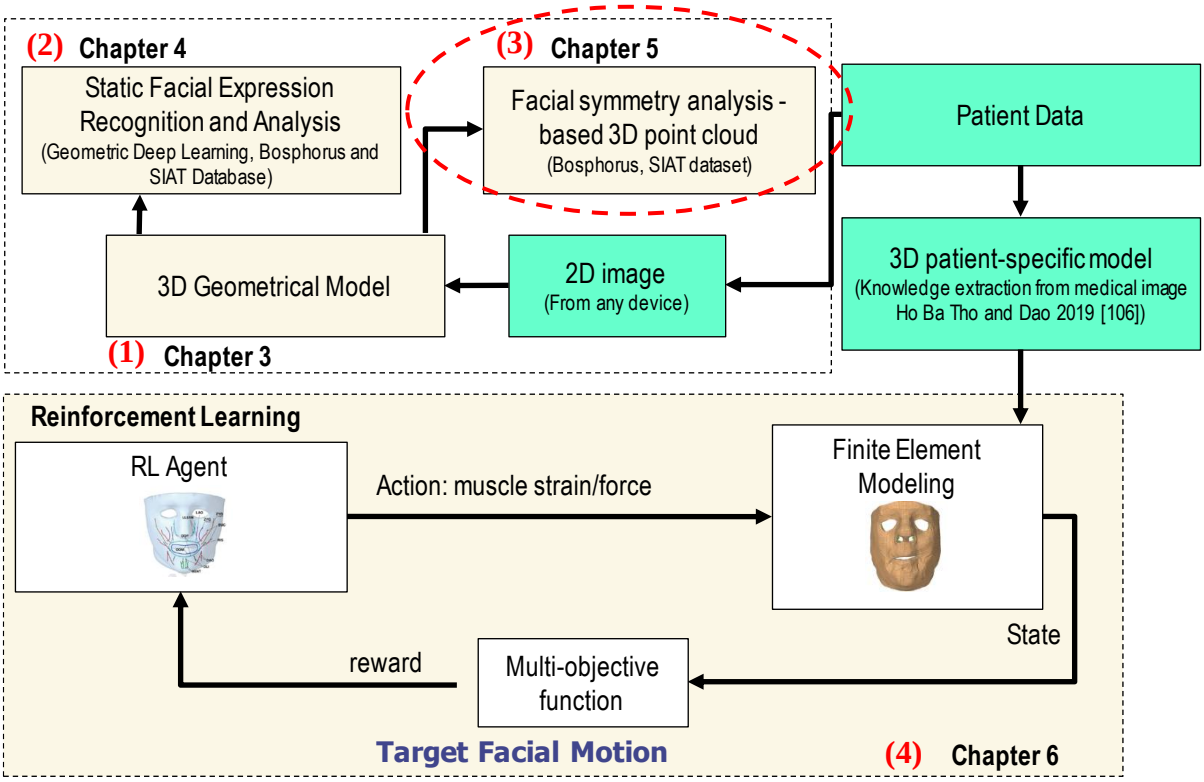


Figure 45. The thesis workflow: the part corresponds with facial symmetry analysis base on 3D point cloud data

5.1. Materials and methods

5.1.1. Learning Databases

Two different 3D facial databases were used for facial symmetry analysis. The first database namely the Bosphorus database [113] acquired the face of 105 subjects (61 males and 44 females) using Inspeck Mega Capturor II 3D devices. In fact, most of the subjects are Caucasian at the ages range from 25 to 35 years old. Each subject was captured a total of 31 to 54 times in different poses and expressions.

SIAT-3DFE database was also used to extract shape descriptors [201]. Two structured light scanners (called CASZM-MVS600) were used to capture the face with high resolution. 500 Asians aged from 18 to 50 with a total 8000 of facial scans were captured. Each subject conducted 16 trials with four basic expressions including happiness, neutral, sadness, and surprise.

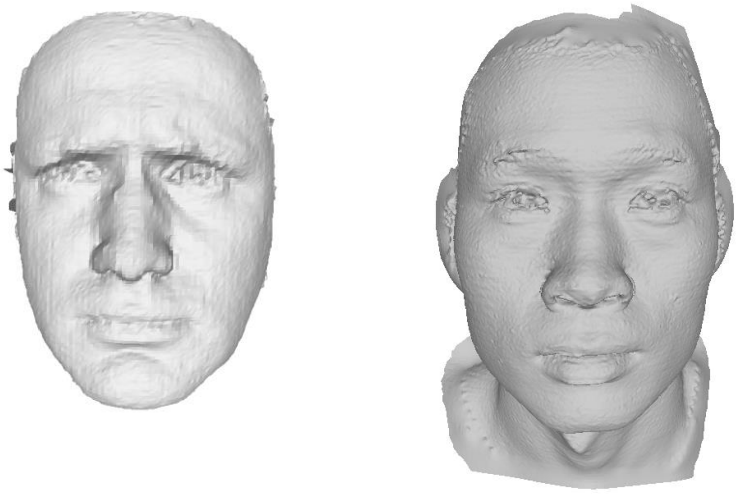


Figure 46. 3D face scans from the Bosphorus database (left) and the SIAT-3DFE database (right).

5.1.2. Data pre-processing procedures

65 subjects including 455 original face scans with seven basic expressions have been chosen for processing from the acquired Bosphorus database. 70 subjects (35 males and 35 females) from the SIAT-3DFE were randomly chosen. The outliner point was initially removed and the noise was reduced using statistical techniques [371]. We cropped the face into the center region for analyzing the face only instead of other regions such as the ear and hair.

The face then was split into 6 parts such as right and left eyes, right and left nose and cheek, right and left mouth, and jaw as in Figure 47. This was done by vertical and horizontal lines go through 3 red points (as shown in Figure 47), which were manually picking. Each part was then normalized by translating with the center of the part at the origin of the coordinate and the coordinate value ranged from -1 to 1.

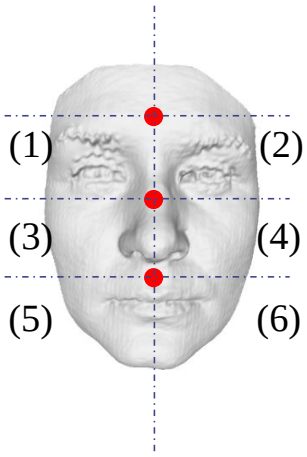


Figure 47. The face was divided into 6 separated parts including left and right eyes, left and right noses, left and right mouth

5.1.3. Descriptors-based geometric deep learning

The 3D facial descriptor analysis was extracted by a well-known geometric deep learning with a hierarchical neural network model called PointNet++. The descriptors are the learned features extracted from the PointNet++ layer. PointNet++ learned the features using three main layers: 1) *Sampling layer* picks a set of points from the input set, which then defines the centroids of this set of points and treats them as local regions; 2) *Grouping layer* creates local regions by searching neighboring points around defined centroids; and 3) *PointNet layer* defines feature vectors by encoding these local regions. These feature vectors are then treated as novel descriptors from the original input point cloud data to perform symmetry analysis.

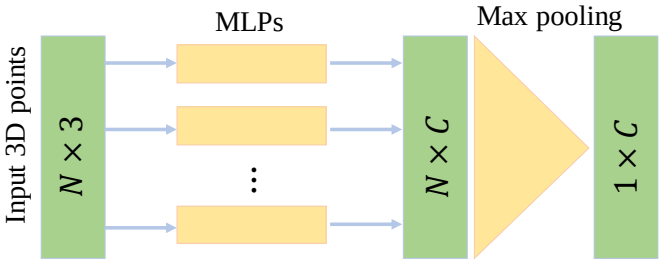


Figure 48. Feature extraction from PointNet, where N represents the pre-defined number of input points, C represents the pre-defined number of learning features, and MLPs represents for the Multi-Layer Perceptron.

PointNet adopts symmetry functions as the multi-layer perceptron to take *n* vectors as input to synthesize the information and generate a new presentation vector that is invariant to the input vector order. Specifically, from an unordered input point set $\{x_1, x_2, \dots, x_n\}$ (where $x_i \in \mathbb{R}^d$), PointNet uses a set function f to convert an input point set into a new feature vector.

$$f(x_1, x_2, \dots, x_n) = \gamma\left(\max_{i=1,\dots,n}\{h(x_i)\}\right)$$

where γ and h present multi-layer perceptron networks (MLP). PointNet independently applies the MLP layer for each input point for learning features and then extracts global features using a max-pooling layer.

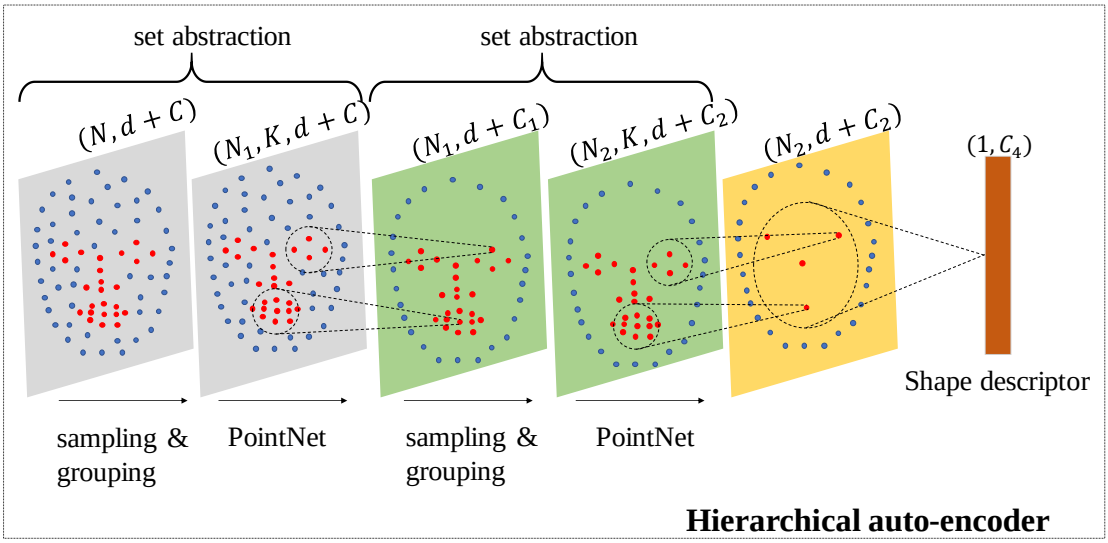


Figure 49. A general hierarchical local feature learning was demonstrated in 2D space. N, N_1, N_2 represent the number of points in d dimensional space and C, C_1, C_2 dimensional feature for each step of applying 3 layers including sampling, grouping, and PointNet. K represents the number of points within a group with centroid points and a pre-set radius.

PointNet learns features for each point in local regions in an independent way, thus in these local regions, the network cannot generalize the local structural information between neighborhood points. PointNet++ resolves this issue by adding two layers namely the sampling and grouping layer to create local regions using the neighborhood of each centroid point. Then the local feature is learned by applying the PointNet layer for these local regions instead of from each individual point.

The model was trained to extract local features on Collaboratory, which is a free research tool. The virtual machine environment with a configuration of a 12GB NVIDIA Tesla K80 GPU, installed CUDA version 10.1, the chosen TensorFlow version 1.15.2, and Python version 3.7.

After feature extraction using a sublayer of PointNet++, each face scan was presented by a 1024-dimensional learned feature vector. These feature vectors are then concatenated to build a matrix with each column is one feature vector presents for each part of the face. Then a principal component analysis was applied for reducing the dimensional space from 1024 to 3 components and still keeping the large variance as shown in Figure 50.

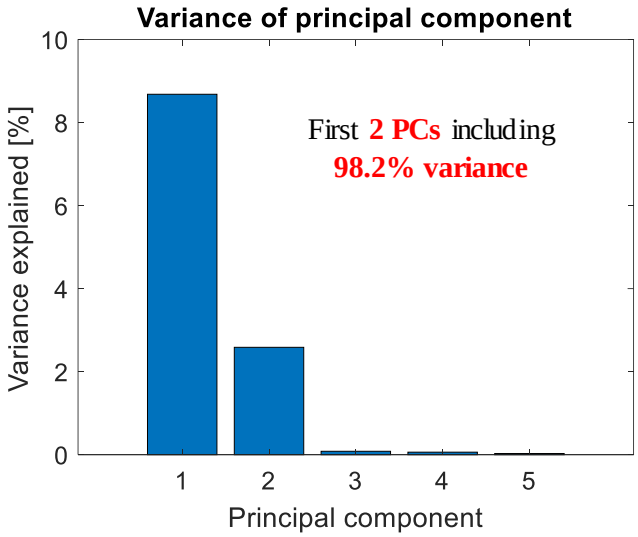


Figure 50. The variance explained by principal component analysis

5.1.4. Descriptor-based PointPCA

The PointPCA method [385] was used to extract the descriptors. After the PCA, a set of low-level geometrical descriptors for each portion of the face were defined and estimated for each point in the input set. Firstly, a set of points as support or local region (with the centroid $\bar{\boldsymbol{p}}_i$) around a point ($\boldsymbol{p}_i$) was determined by r-research for each point in coordinate ($\boldsymbol{u}_x, \boldsymbol{u}_y, \boldsymbol{u}_z$). Secondly, the covariance matrix is computed and applied PCA for this set. Finally, computed eigenvalues ($\lambda_1 > \lambda_2 > \lambda_3$) and eigenvectors ($\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3$) were used as the descriptors to encode the local 3D shape properties. Other descriptors based on PointPCA are defined in *Table 6*. The first 3 descriptors are used for further analysis to test the feasibility of the research, while the remaining descriptors will be used the future work.

Table 6. definition of the 3D shape descriptors

Descriptor	Definition
Eigenvalues	$d_v = \lambda_v, v \in \{1, 2, 3\}$
Sum of eigenvectors	$d_4 = \sum_v \lambda_v$
Linearity	$d_5 = (\lambda_1 - \lambda_2)/\lambda_1$
Planarity	$d_6 = (\lambda_2 - \lambda_3)/\lambda_1$
Sphericity	$d_7 = \lambda_3/\lambda_1$
Anisotropy	$d_8 = (\lambda_1 - \lambda_3)/\lambda_1$
Omni-variance	$d_9 = \sqrt[3]{\lambda_1 \cdot \lambda_2 \cdot \lambda_3}$
Eigen-entropy	$d_{10} = -\sum_v \lambda_v \cdot \ln(\lambda_v)$
Surface variation	$d_{11} = \lambda_3 / \sum_v \lambda_v$
Roughness	$d_{12} = \lvert(\boldsymbol{p}_i - \bar{\boldsymbol{p}}_i) \cdot \boldsymbol{e}_3\rvert$
Parallelity x	$d_{13} = 1 - \lvert\boldsymbol{u}_x \cdot \boldsymbol{e}_3\rvert$
Parallelity y	$d_{14} = 1 - \lvert\boldsymbol{u}_y \cdot \boldsymbol{e}_3\rvert$
Parallelity z	$d_{15} = 1 - \lvert\boldsymbol{u}_z \cdot \boldsymbol{e}_3\rvert$

5.2. Facial symmetry analysis results

The objective is to quantify whether the technique is able to analyze the symmetry or asymmetry of the face based on descriptors extracted by two techniques: PointNet++ and PointPCA. Figure 51 displays descriptors extracted by the PointNet++ layer for the eye on the left (red points) and right (cyan points) in the same population including Caucasian (a) and Asian (b). The discrimination between the left and right eyes is more observable for eye descriptors of Caucasians, while these descriptors of Asians are overlaid between two sides.

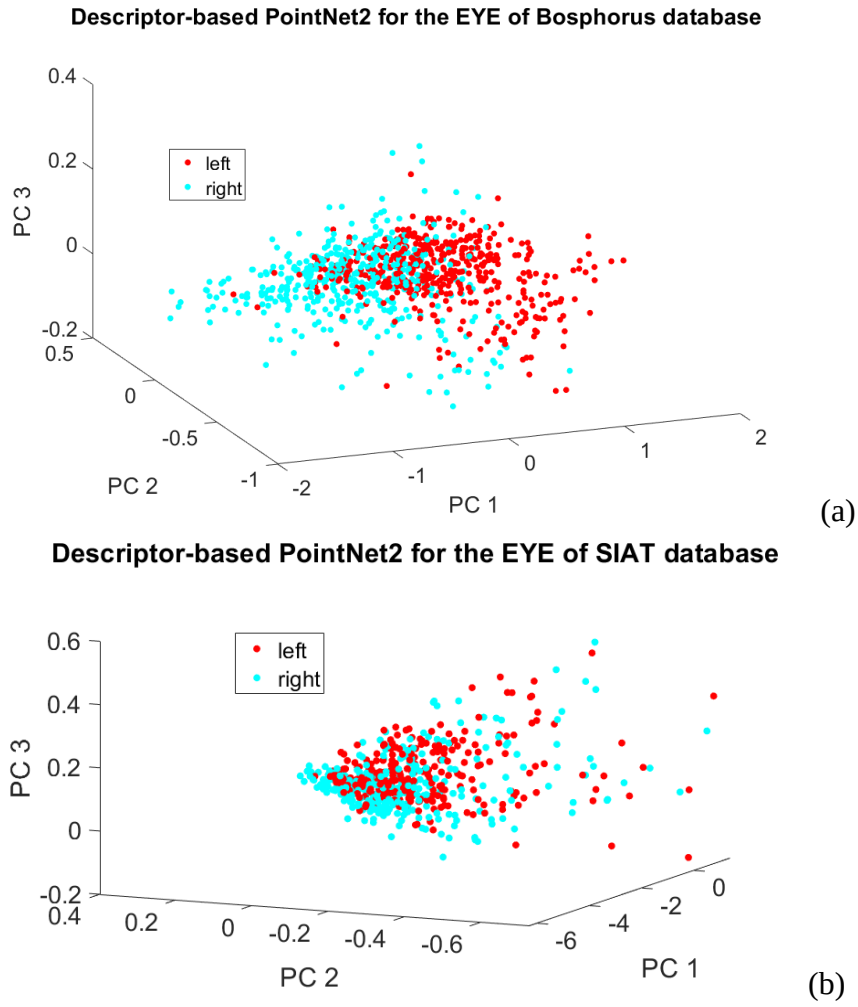


Figure 51. Discriminative learned descriptors extracted by PointNet++ for symmetry analysis at the eye region for two populations including Caucasian (a) and Asian (b)

Descriptors extracted by PointNet++ layer for the nose region on the left (red points) and right (cyan points) were illustrated in Figure 52 in the same population including Caucasian (a) and Asian (b). A similar pattern was observed in Figure 51. In this region, the discrimination between the left and right noses is more observable in Caucasians compared to Asians, based on the learned descriptors extracted by the PointNet++ layer.

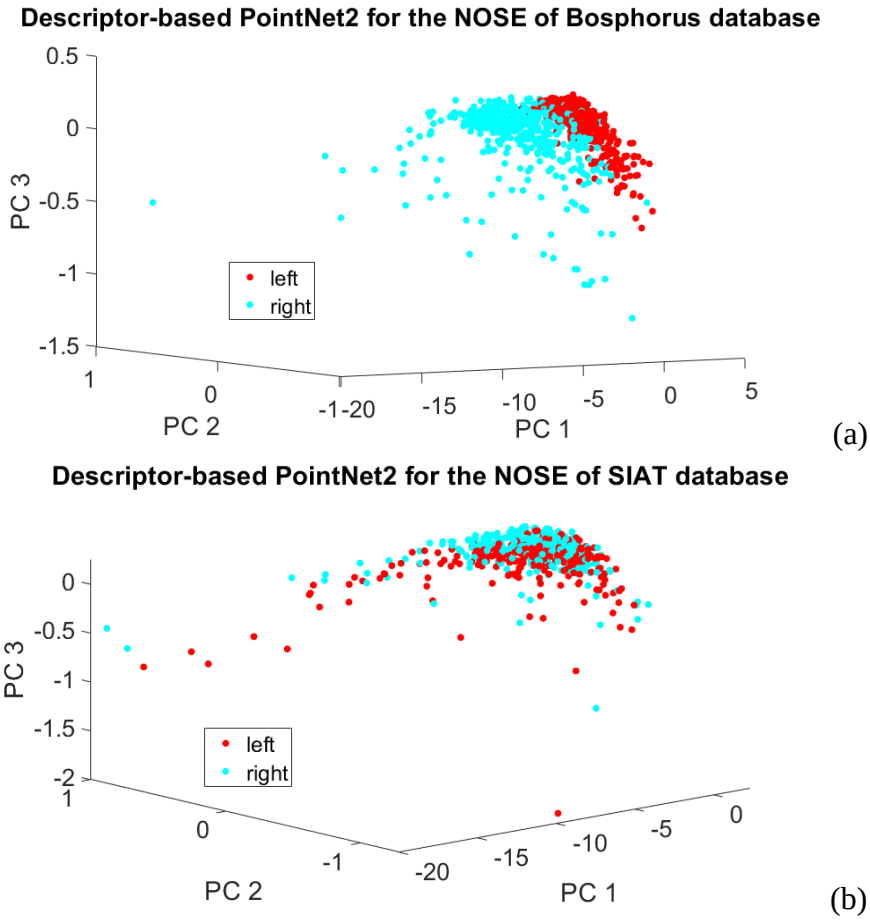


Figure 52. Discriminative learned descriptors extracted by PointNet++ for symmetry analysis at the nose region for two populations including Caucasian (a) and Asian (b)

Figure 53 shows a similar pattern as in Figure 52 and Figure 51 for the mouth region. The discrimination between the left mouth and right mouth is more observable in Caucasians compared to Asians, based on the learned descriptors extracted by the PointNet++ layer

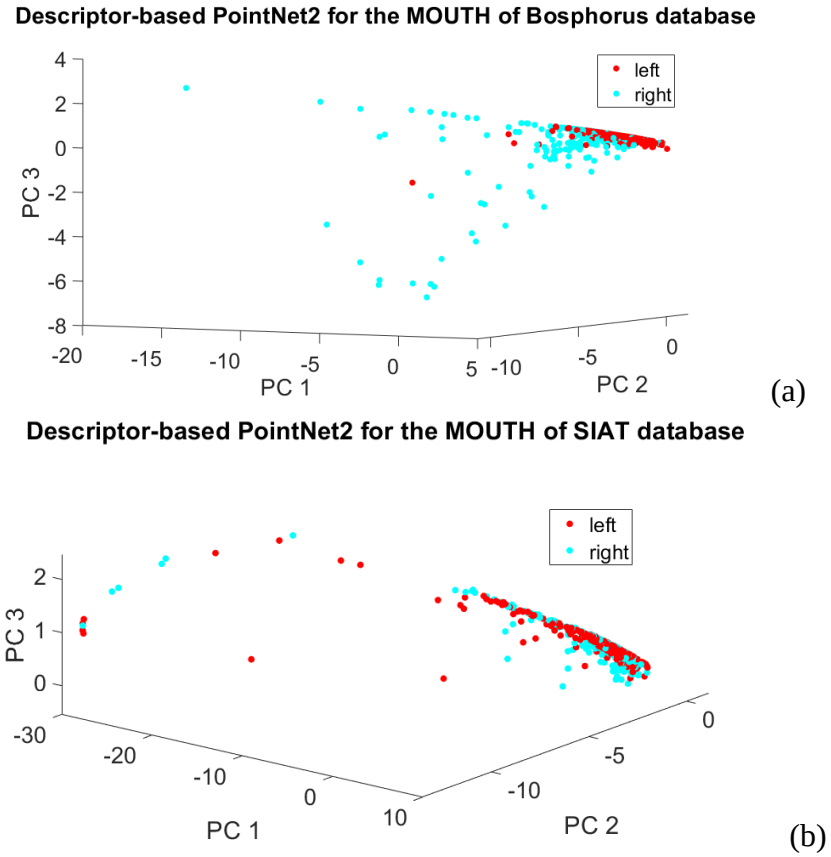
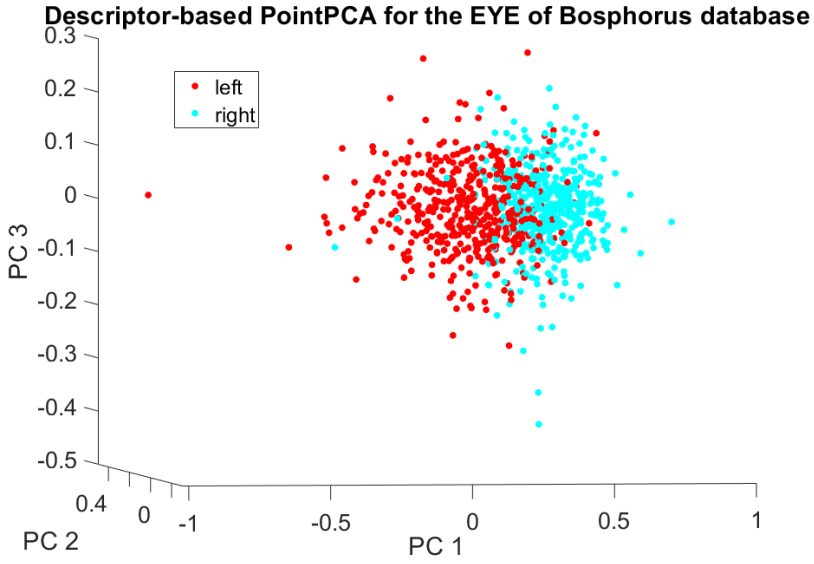
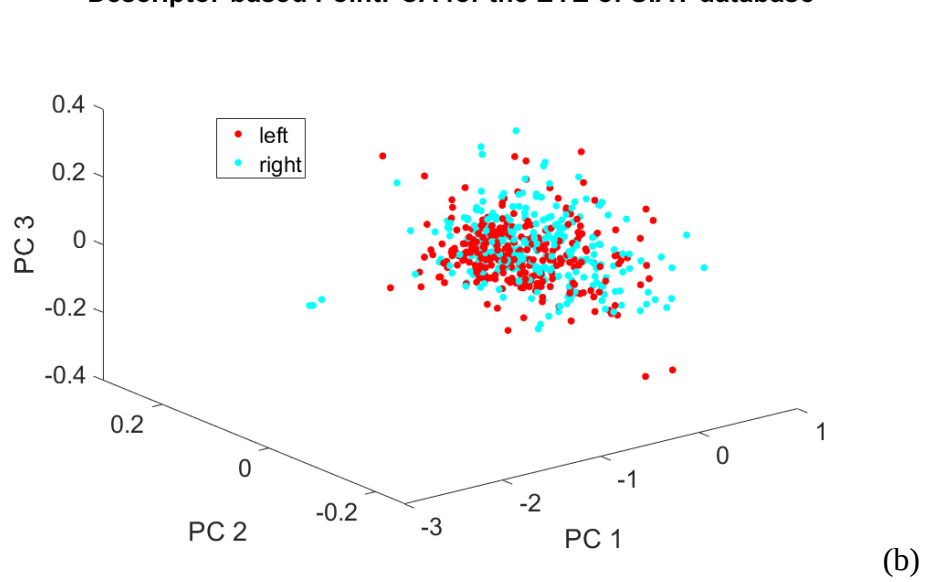


Figure 53. Discriminative learned descriptors extracted by PointNet++ for symmetry analysis at the mouth region for two populations including Caucasian (a) and Asian (b)

Similarly, descriptors extracted by PointPCA were shown in Figure 54, Figure 55, and Figure 56 for all three regions of eye, nose, and mouth on the left (red points) and right (cyan points) in the same population including Caucasians (a) and Asians (b). The observed result shows that the discrimination between the left eye and right eye is more observable in Caucasians (Bosphorus database) compared to in Asians (SIAT database), based on the learned descriptors extracted by PointPCA.



(a)



(b)

Figure 54. Discriminative hand-crafted descriptors-based PointPCA for symmetry analysis at the eye region for two populations including Caucasian (a) and Asian (b)

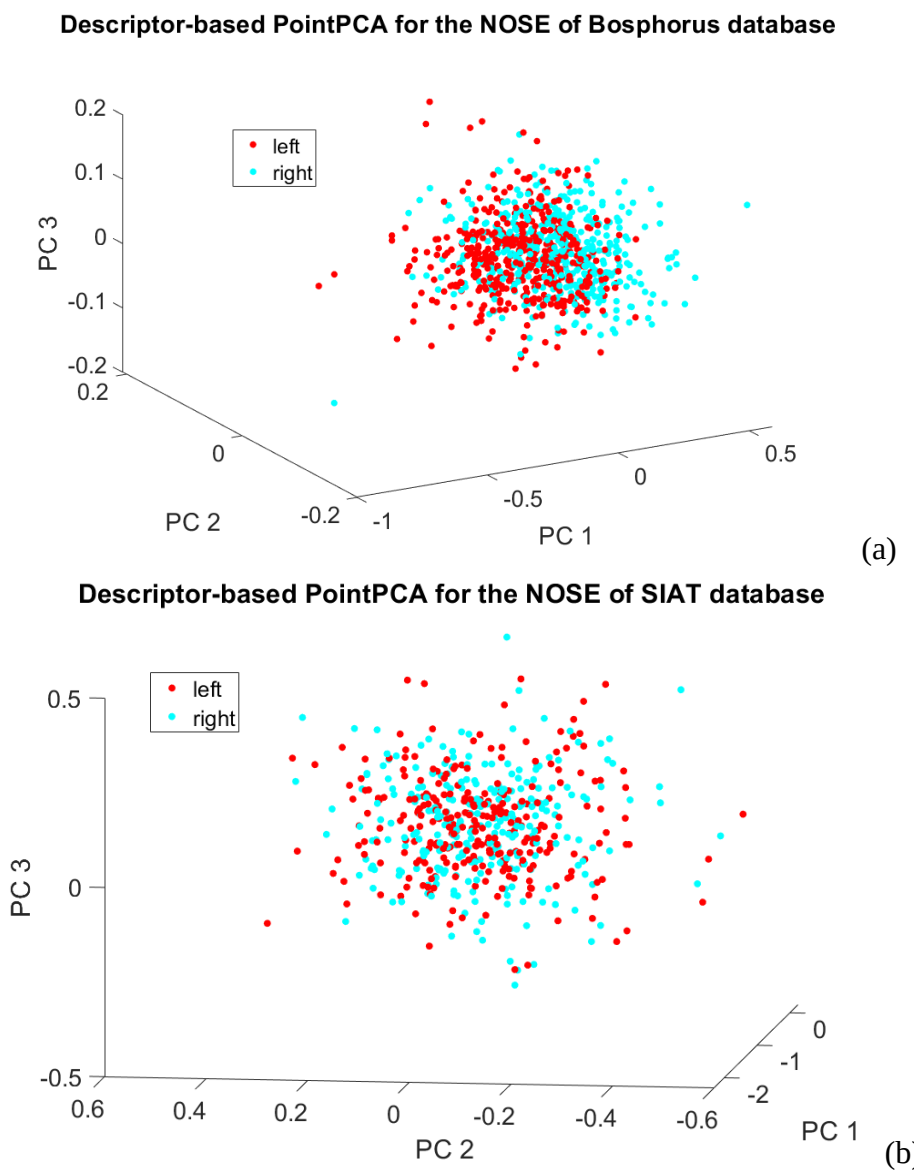
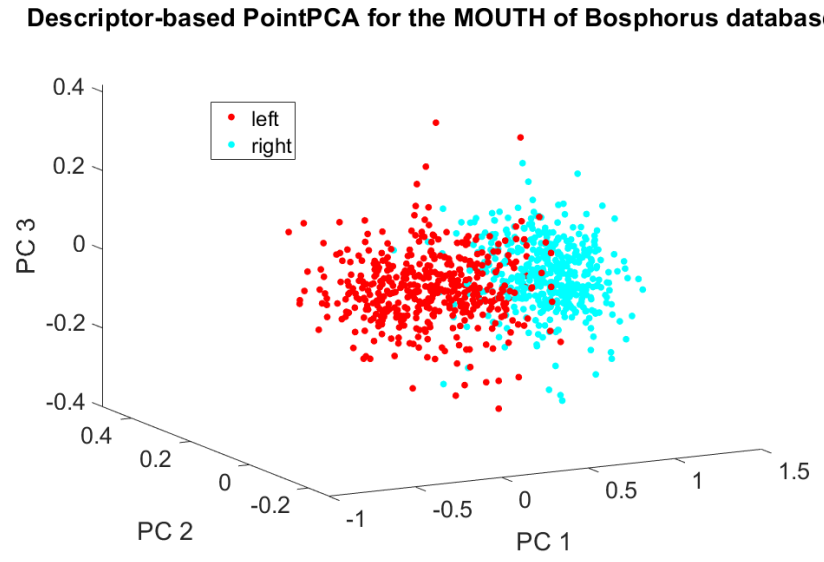
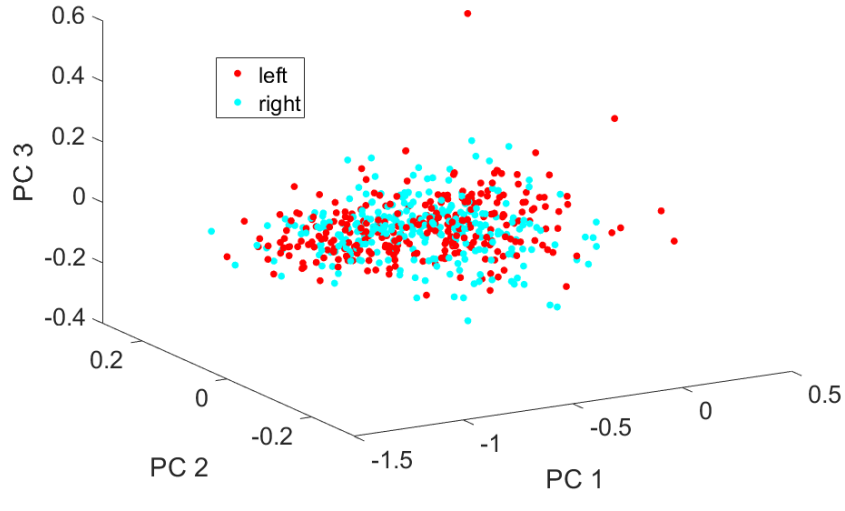


Figure 55. Discriminative hand-crafted descriptors-based PointPCA for symmetry analysis at the nose region for two populations including Caucasian (a) and Asian (b)



(a)



(b)

Figure 56. Discriminative hand-crafted descriptors-based PointPCA for symmetry analysis at the mouth region for two populations including Caucasian (a) and Asian (b)

The discriminative property was shown in Figure 57 based on learned descriptors between two populations including Caucasians (red points) and Asians (green points). These descriptors were extracted by PointNet++ and PointPCA in three different regions such as eyes, nose, and mouth. The figure shows that the discrimination of two populations is more observable in descriptors extracted by PointNet++ (left column) that represent all three regions eye, nose, and mouth compared to that of PointPCA (right column).

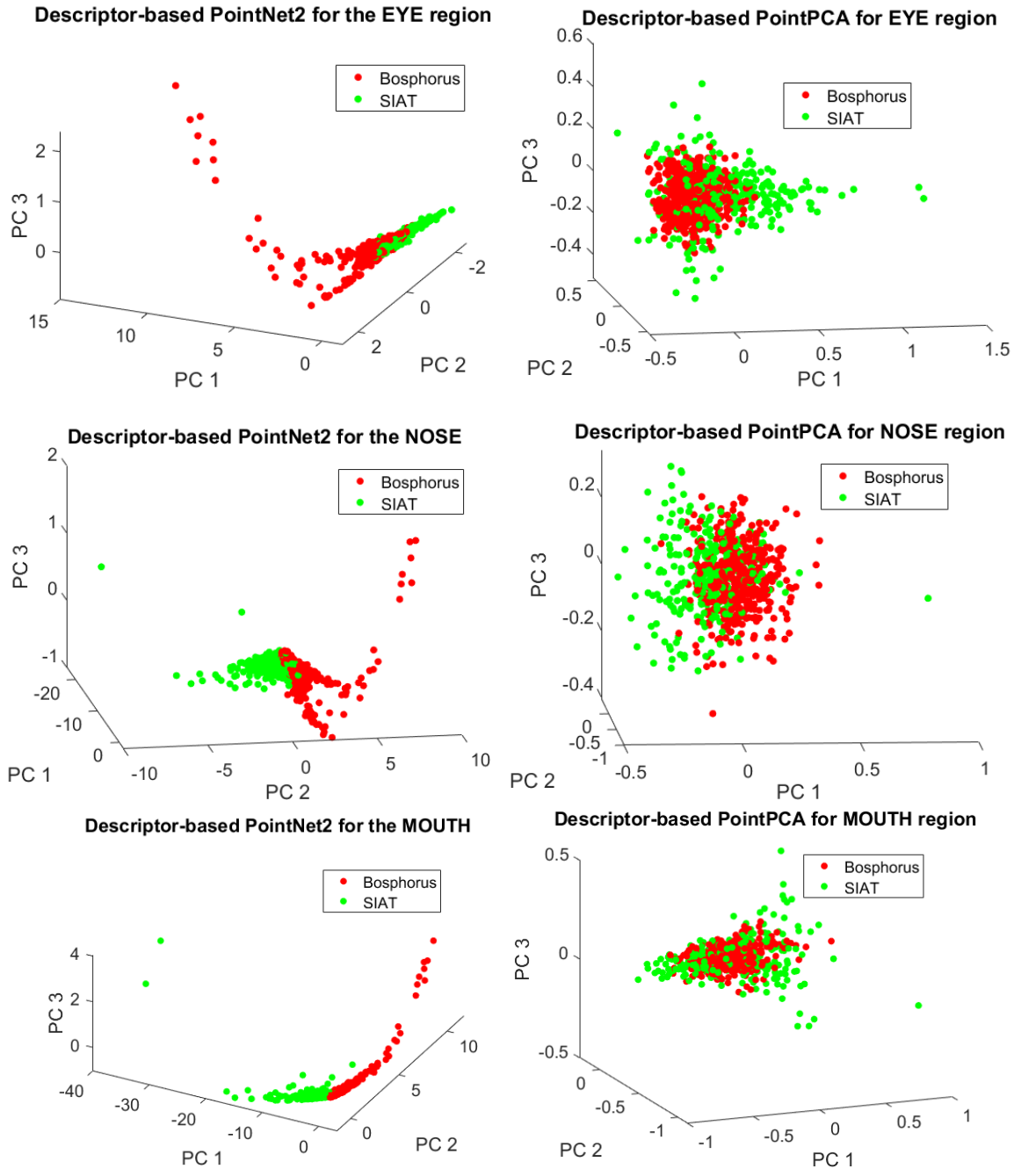


Figure 57. Discriminative learned descriptors for the eye, nose, and mouth regions between two populations including Caucasian (the Bosphorus database) and Asian (the SIAT database)

The separation between the left and right sides of the face in the same population (Caucasians or Asians) was illustrated in Table 7 and Table 8. This is based on descriptors

extracted by both PointNet++ and PointPCA for three different regions such as eye, nose, and mouth. In the eye region, descriptors extracted by PointNet++ for Caucasians show a statistically significant difference between the left and right sides in the first principal axis, while these descriptors of Asians indicate a statistically significant difference in the first and the third principal axes. In the nose region, descriptors of Asians show a statistically significant difference in all three principal axes, while only the first principal axis of that in Caucasians shows a statistically significant difference. In contrast, the mouth region of Caucasians shows significant differences in all three principal axes, while that of Asians there is no statistically significant difference in all principal axes.

Table 7. The symmetry of the face property between left and right using descriptor-based PointNet++

Comparison		Descriptor value in PC1	Descriptor value in PC2	Descriptor value in PC3
Caucasian	Left eye	0.47±0.46	0.005±0.18	-0.02±0.05
	Right eye	-0.24±0.52	0.004±0.16	-0.02±0.07
	p value	< 0.05	> 0.05	> 0.05
	Left nose	1.7±0.45	-0.12±0.13	0.03±0.13
	Right nose	-0.86±1.77	0.08±0.22	-0.05±0.21
	p value	< 0.05	> 0.05	> 0.05
	Left mouth	1.55±0.96	0.51±0.48	-0.06±0.17
	Right mouth	0.38±1.92	-0.28±1.82	-0.16±0.91
	p value	< 0.05	< 0.05	< 0.05
Asian	Left eye	-0.022±0.7	-0.015±0.17	0.043±0.09
	Right eye	-0.35±0.95	-0.0001±0.2	0.022±0.09
	p value	< 0.05	> 0.05	< 0.05
	Left nose	-1.09±4.78	-0.007±0.5	-0.02±0.29
	Right nose	-0.27±3.01	0.07±0.22	0.05± 0.17
	p value	< 0.05	< 0.05	< 0.05
	Left mouth	-0.65±4.58	-0.4±2.11	0.22±0.41
	Right mouth	-0.62±4.39	-0.28±1.79	0.19±0.42
	p value	> 0.05	> 0.05	> 0.05

Descriptors extracted by PointPCA show that there were only statistically significant differences between the left and right sides in the first principal axis of all regions for Caucasians and the eye and the nose regions for Asians.

Table 8. The symmetry of the face property between left and right using descriptor-based PointPCA

Comparison		Descriptor value in PC1	Descriptor value in PC2	Descriptor value in PC3
Caucasian	Left eye	0.09±0.176	-0.002±0.08	0.002±0.08
	Right eye	0.38±0.11	0.003±0.08	0.0006±0.08
	p value	< 0.05	> 0.05	> 0.05
	Left nose	0.08±0.15	0.005±0.05	-0.001±0.05
	Right nose	0.2±0.15	-0.0008±0.1	0.002±0.05
	p value	< 0.05	> 0.05	> 0.05
	Left mouth	0.05±0.278	0.005±0.083	-0.004±0.08
	Right mouth	0.7±0.17	-0.002±0.08	0.001±0.08
	p value	< 0.05	> 0.05	> 0.05
Asian	Left eye	-0.44±0.35	0.002±0.07	-0.008±0.08
	Right eye	-0.33±0.39	-0.004±0.07	0.004±0.08
	p value	< 0.05	> 0.05	> 0.05
	Left nose	-0.277±0.19	-0.004±0.14	0.002±0.13
	Right nose	-0.17±0.17	-0.002±0.13	-0.004±0.13
	p value	< 0.05	> 0.05	> 0.05
	Left mouth	-0.62±0.34	-0.002±0.08	0.006±0.08
	Right mouth	-0.62±0.289	-0.002±0.08	-0.0004±0.1
	p value	> 0.05	> 0.05	> 0.05

Discriminative properties between two populations (Caucasians and Asians) were illustrated in Table 9 and Table 10 based on descriptors extracted by both PointNet++ and PointPCA for all three regions eye, nose, and mouth. The tables show that there are statistically significant differences between the descriptors of the face of Caucasians and Asians (p-value < 0.05) in all 3 principal axes for the descriptors extracted by PointNet++ layers, while there is, only in the first principal axis, descriptors extracted by PointPCA have p-value < 0.05.

Table 9. Discriminative property of the descriptors-based PointNet++ between two populations

Comparison		Descriptor value in PC1	Descriptor value in PC2	Descriptor value in PC3
Eye region	Caucasian	0.42±1.94	0.25±0.53	-0.03±0.27
	Asian	-0.68±0.3	-0.41±0.6	0.06±0.12
	p value	< 0.05	< 0.05	< 0.05
Nose region	Caucasian	0.39±1.47	0.72±1.36	-0.03±0.28
	Asian	-0.63±2.53	-1.18±0.91	0.05±0.18
	p value	< 0.05	< 0.05	< 0.05
Mouth region	Caucasian	1.25±0.8	0.47±1.84	0.02±0.23
	Asian	-2.04±3.9	-0.76±0.62	-0.04±0.36
	p value	< 0.05	< 0.05	< 0.05

Table 10. Discriminative property of the descriptors-based PointPCA between two populations

Comparison		Descriptor value in PC1	Descriptor value in PC2	Descriptor value in PC3
Eye region	Caucasian	-0.07±0.11	0.0007±0.09	0.0007±0.09
	Asian	0.1135±0.25	-0.0012±0.1	-0.0011±0.1
	p value	< 0.05	> 0.05	> 0.05
Nose region	Caucasian	0.05±0.1	0.001±0.09	-0.002±0.1
	Asian	-0.08±0.14	-0.001±0.12	0.004±0.11
	p value	< 0.05	> 0.05	> 0.05
Mouth region	Caucasian	-0.018±0.16	-0.001±0.04	0.005±0.05
	Asian	0.03±0.29	0.002±0.1	-0.007±0.09
	p value	< 0.05	> 0.05	> 0.05

5.3. Discussion and conclusion

Facial symmetry is important in recognizing the impression of beauty and attraction. Facial palsy patients and patients with facial transplantation are more likely to have an asymmetric face. Thus, analyzing facial symmetry is important to support the correct assessment and diagnosis of facial palsy and facial transplantation patients. This is of helps the doctor to design efficient individual rehabilitation programs. Traditional methods were either subjectively conducted based on several facial grading systems or based on 2D image

processing, which requires hand-crafted feature engineering. Thus, the present chapter proposed to use an end-to-end deep learning solution for presenting the face or part of the face into a novel learned descriptor, which then can be used as an indicator for analyzing facial symmetry.

The present chapter proposed to use a new class of deep learning networks called geometric deep learning with PointNet++ model and another hand-crafted feature engineering method called PointPCA. Geometric deep learning allows exploring more complex types of data as non-Euclidean data, which can be applied to learn directly from 3D point cloud data. This offers to analyze the 3D face scan to overcome the challenging problem of extreme pose variations and 2D image variations (brightness, pose) for 2D image processing. Two databases have been used including the Bosphorus database (Caucasian subjects) and SIAT (Asian subjects). The purpose is to test whether the technique is able to analyze the face in terms of facial symmetry or asymmetry. And in the meantime, the method also investigated the discrimination between the face of Caucasians and Asians. The obtained results show that there is a statistically significant difference in the descriptors representing facial symmetry of Caucasians' mouth and Asian's nose. Furthermore, there were also statistically significant difference between descriptors representing the face of Caucasians and Asians. The result also shows that descriptors extracted by PointNet++ show a more discriminative property than that of PointPCA.

Despite the fact that PointNet++ can operate directly on 3D point clouds, 3D raw point clouds still require preprocessing operations such as noise removal, face cropping in the central region, and data normalization. Moreover, even though analyzing the face based on **learned descriptors** one can provide **an indicator for symmetry or asymmetry**. But it cannot measure the level of the symmetry of the face. In order to do that, the developed model will be applied to a database including facial palsy patients to provide an **indicator for grading the asymmetry of the face of facial palsy patients**. This has to collaborate with clinicians.

The outcome of this chapter is a paper in preparation: "*Facial Symmetry Analysis based on Novel Shape Descriptors between Two Different Populations*" to be submitted to *Medical & Biological Engineering & Computing (Q2, IF@2020 = 2.602).*

Chapter 6:

Reinforcement Learning coupled with Finite Element Modeling for Facial Motion Learning

Facial palsy patients or patients with facial transplantation have abnormal facial motion due to altered facial muscle functions and nerve damage. Moreover, involved patients also suffer asymmetric face effect, which indicates the imbalance and inequality of facial structure in terms of shape, size, location, and arrangement of left and right components on the sagittal plane. In fact, the recovery of a symmetric face with balanced functionalities requires a complex rehabilitation process in which patients must practice patient-specific facial movements. It is important to note that current facial rehabilitation has mainly been based on a mirror approach to monitor the visual qualitative feedback from the rehabilitation exercise. This strategy is ineffective and subjective without any feedback. Computer-aided systems and physics-based models have been developed to provide objective and quantitative information. However, the predictive capacity of these solutions is still limited to explore the facial motion patterns with emerging properties. The present study aims to couple reinforcement learning and finite element modeling for facial motion learning and prediction. The main objective is to provide, for the first time, the modeling workflow for this complex coupling and then to evaluate different learning strategies to establish motion patterns of the face during facial expression motions. Our novel solution will explore the patient-specific facial motions without a priori data from the patient and then provides a set of facial muscle activation and coordination patterns for a specific rehabilitation-oriented movement (e.g. symmetry or smile).

A novel modeling workflow for learning facial motion was developed. The method aims to use a subject-specific model of the face. However, developing a subject-specific model was a complex task. Moreover, an exchange information protocol between two different platforms (reinforcement learning and simulation environment) was also complex. Thus, in the first step, an existing physically-based model of the face within the Artisynth modeling platform was used to test the feasibility of the method. This model has been published and well developed for facial mimic simulation. Information exchange protocol was proposed to exchange capacity between reinforcement learning and rigid multi-bodies dynamics simulation. Two reinforcement learning algorithms (deep deterministic policy gradient (DDPG) and Twin-delayed DDPG (TD3)) were used and implemented to drive the simulations of symmetry-oriented and smile movements. Numerical outcomes were compared to experimental observations (Bosphorus database) for evaluation and validation purposes.

The part corresponds with task number 4 in the workflow (Figure 58).

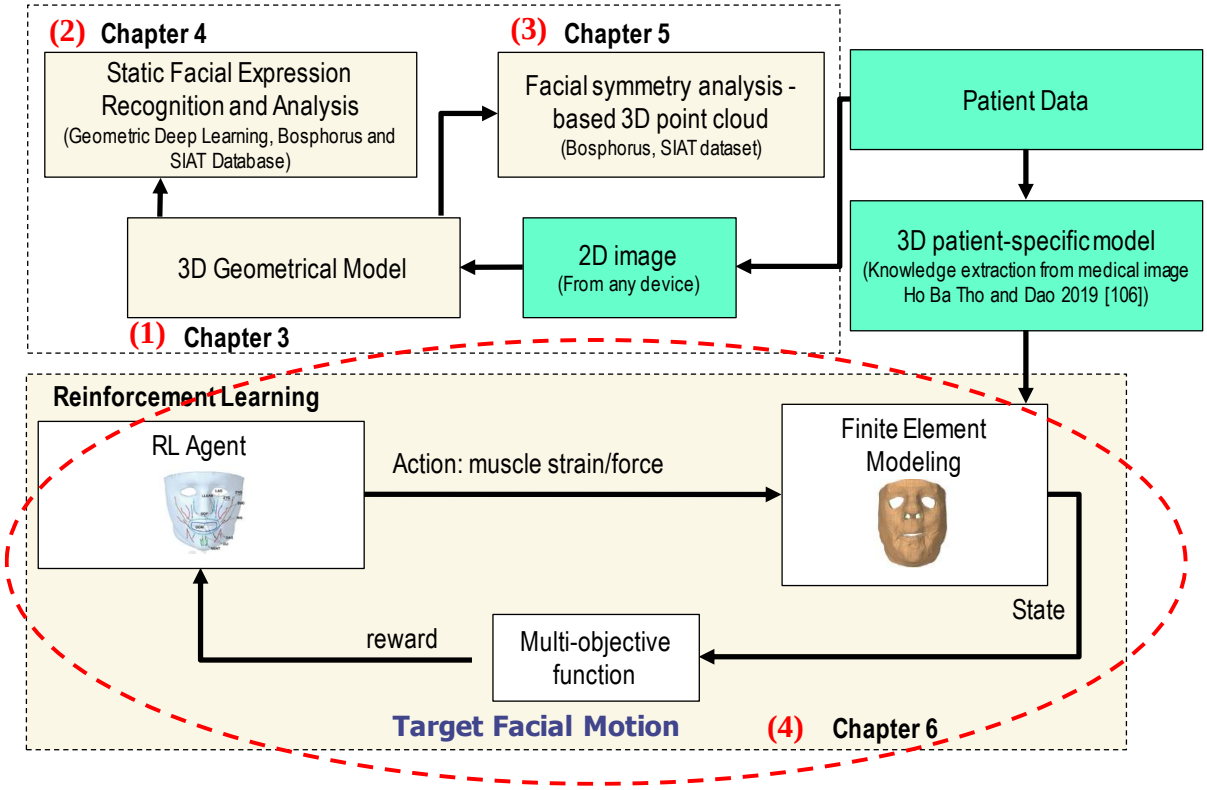


Figure 58. Reinforcement learning coupled with finite element model of the face for facial motion learning.

6.1. Materials and methods

6.1.1. Novel coupling workflow between reinforcement learning and finite element modelling

Our novel simulation workflow requires two main components (Figure 59): 1) a reinforcement learning agent (a human face) having a policy that decides what action (muscle excitations) to take when it observes a state (facial motion) and 2) a finite element modeling and simulation environment. The coupling between the finite element simulation environment and the reinforcement learning process is managed by an information exchange protocol. More precisely, in the beginning, the reinforcement learning agent observes the state of the face using the positions of selected key points (Figure 60). Secondly, the policy predicts values of muscle excitations, which are then applied to the biomechanical model of the face for a physical simulation. Then, the simulation environment returns the positions of selected key points after simulation. And finally, these positions are used to compute the reward value by pre-designed multi-objective functions (related to symmetry or smile exercises), which is then used to update training parameters for the training process.

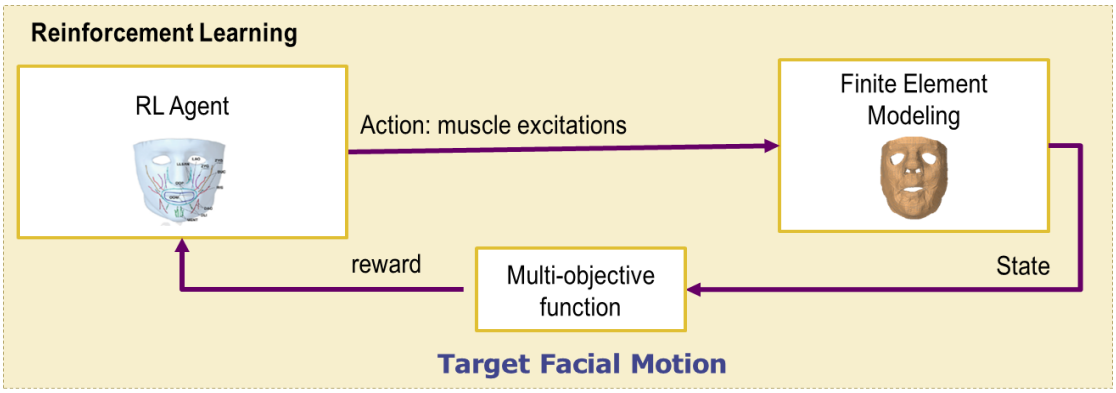


Figure 59. Overview of the novel coupling workflow between reinforcement learning and finite element modeling

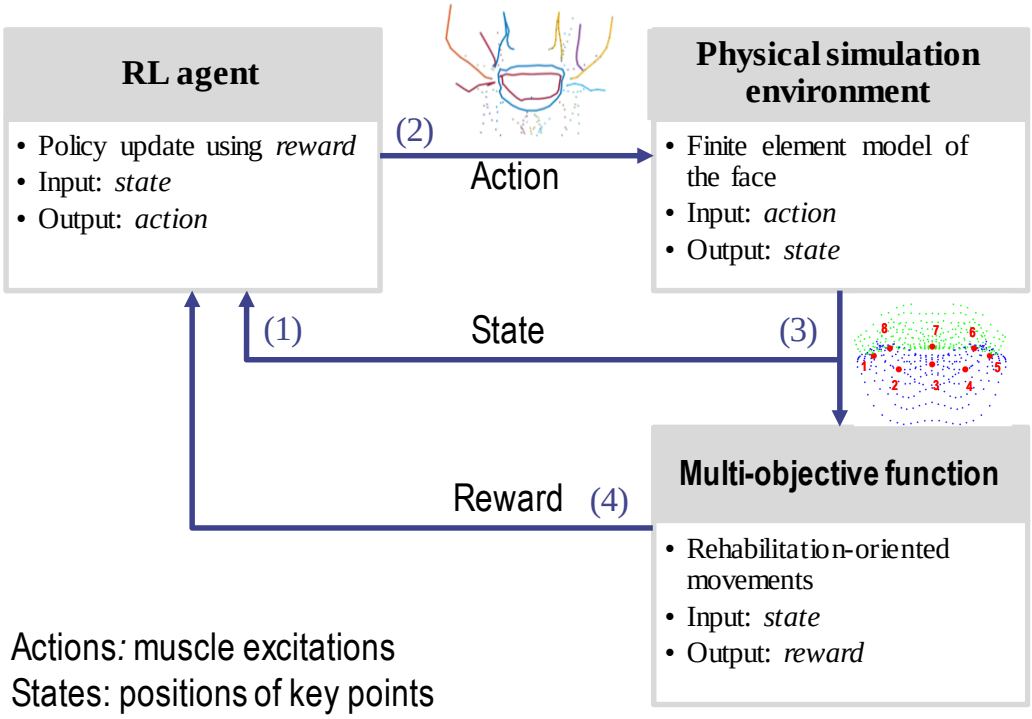


Figure 60. Detailed flowchart of the interaction between reinforcement learning and finite element modeling processes.

6.1.2. Face finite element model

A physically-based model of the face within the Artisynth modeling platform was used (Figure 61a). This model has been developed from previous research [122], [386]–[388]. The face finite element model includes three components such as 1) a soft-tissue component with the hypodermis, dermis, and epidermis layers, 2) a cranium and maxilla component, and 3) a jaw-hyoid component [389]. Using a 2.6GHz Core 2 Duo CPU, each simulation for the original face model took around 35 minutes [390]. When reinforcement learning often requires thousands or even millions of episodes for training, this amount of computational time for each simulation is excessive. To reduce computational cost and accelerate the training process, the facial model is simplified by keeping only the soft-tissue component with ten orofacial muscles (Levator Anguli Oris (LAO), Levator Labii Superioris Alaeque Nasi (LLSAN), Buccinator (BUC), Zygomaticus (ZYG), Depressor Anguli Oris (DAO), Risorius (RIS), Depressor Labii Inferioris (DLI), Mentalis (MENT), Orbicularis Oris Peripheralis (OOP), Orbicularis Oris Marginalis (OOM)) (Figure 61c). The soft tissue finite element mesh consists of 6342 brick elements (with 6024 hexahedrons and 318 wedges) and 8720 nodes. The activation for the face model results from the orofacial muscle strain and force. Ten orofacial muscles are modeled and attached to the lower face that applies muscle forces in terms of muscle excitations onto the finite element model. Muscle fibers are modeled by a set of uniaxial cable elements. For example, the zygomatic ligaments are represented by fixing all degrees of freedom of soft tissue nodes that are in the region where these ligaments attach to the maxilla. The soft tissue constitutive equation for the hypodermis layer is based on a Mooney-Rivlin constitutive equation, and Fung constitutive equation for the epidermis and dermis layer as in the Flynn et al. paper [122] (2015). The mechanical

characteristics (such as force-displacement response, pre-stress behaviors, non-linear, anisotropic, and viscoelastic constitutive laws) for the skin layer were estimated based on a combination of in vivo experiments and numerical methods. Muscles are modeled as continuous sets of cable elements, which activate in tension as point-to-point Hill-type models and are aligned along element edges. The mechanical property evolution of muscle contraction comprises muscle contractile fibres (active part), muscle body (passive part), and the stress stiffening effect [386]. The movements of the mandible generated by muscles of mastication are not handled yet in the model. Thus, the superficial muscles, which are muscles around the lip region, involved in facial mimics are focused. Two finite element models of the face correspond with the modeling of the symmetric face (Figure 61a) and the asymmetric face (Figure 61b).

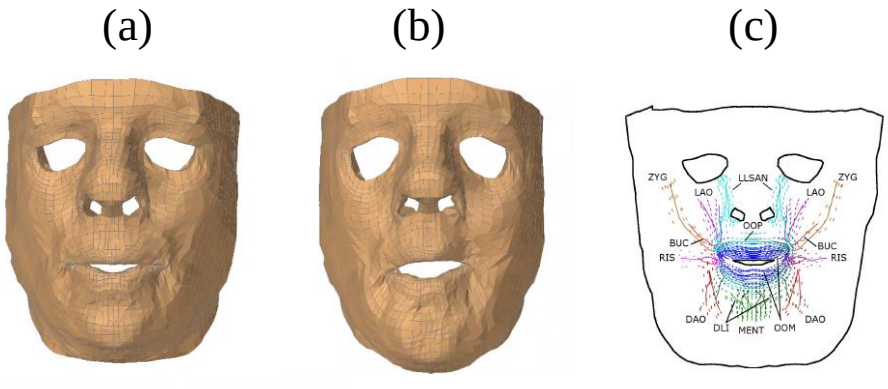


Figure 61. The face finite element model (a) is referred to as the symmetric face, (b) referred to as the asymmetric face (unbalanced deformation between left and right sides), and related facial muscle network (c).

6.1.3. Reinforcement Learning for Facial Motion Control

6.1.3.1. Reinforcement learning model and algorithms

Reinforcement learning (RL) aims to find a policy, $\pi(a|s)$, which maps the state space to the action space and instructs the agent on how to make decisions that maximize the long-term cumulative reward inspired by a reward function $r(s,a)$, where a is the action needs to take in the state s. Bellman equations are solved to find the optimal policy. In this present study, two RL algorithms were used. The first algorithm is the Deep Deterministic Policy Gradient (DDPG) in which the Bellman equation was solved by combining a deep neural network for learning Q function and a deterministic policy gradient algorithm for learning a policy. This is off-policy reinforcement learning used for continuous state and action spaces, which is suitable for our problem. The second used algorithm is the Twin Delayed DDPG (TD3). DDPG is often brittle with the tuning process for hyperparameters. It usually fails when exploiting the error in the Q-function, the learned Q function starts to overestimate Q-values resulting in policy breaking. Twin delayed DDPG copes with this issue and improves performance by applying three tricks: 1) Clipped Double-Q Learning: learning two Q-functions (twin) and using smaller Q-values for the Bellman error loss functions, 2) Delayed

Policy Updates: sparse updating the policy compared to the Q-function, 3) Target Policy Smoothing: adding noise to target actions to reduce exploiting Q-function errors.

The network architecture of DDPG contains the actor network and the critic network. Each network has two hidden layers with 64 nodes (Figure 62). The actor network inputs the state vector while outputs the action vector. The input of the critic network contains both the action vector output from the actor network and the state vector, while the output is the predicted Q-value. TD3 has the same architecture as in DDPG except it has two critic networks (Figure 63).

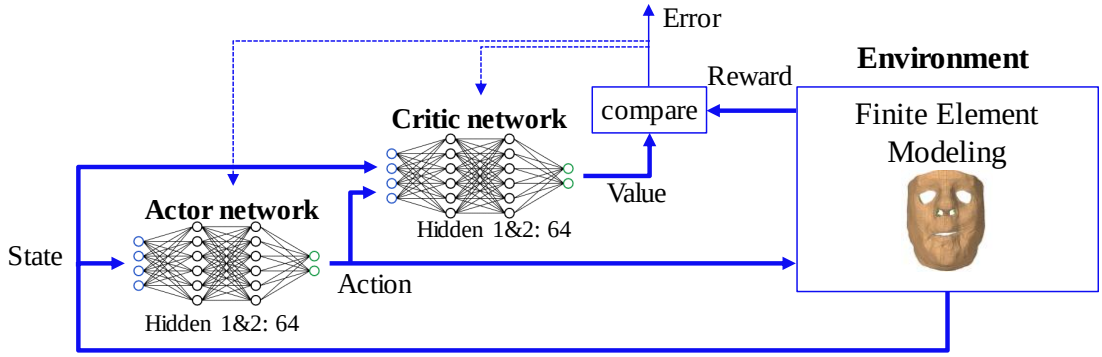


Figure 62. The network architecture of DDPG: one actor network and one critic network.

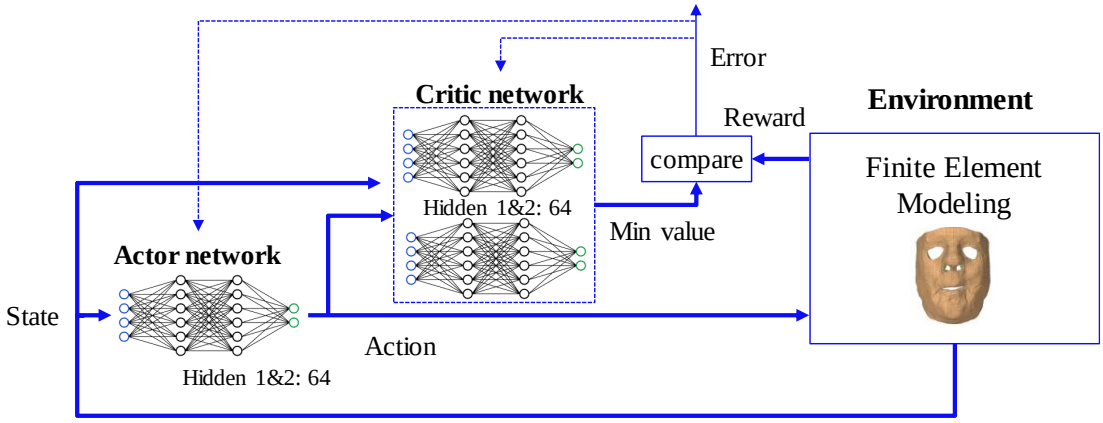
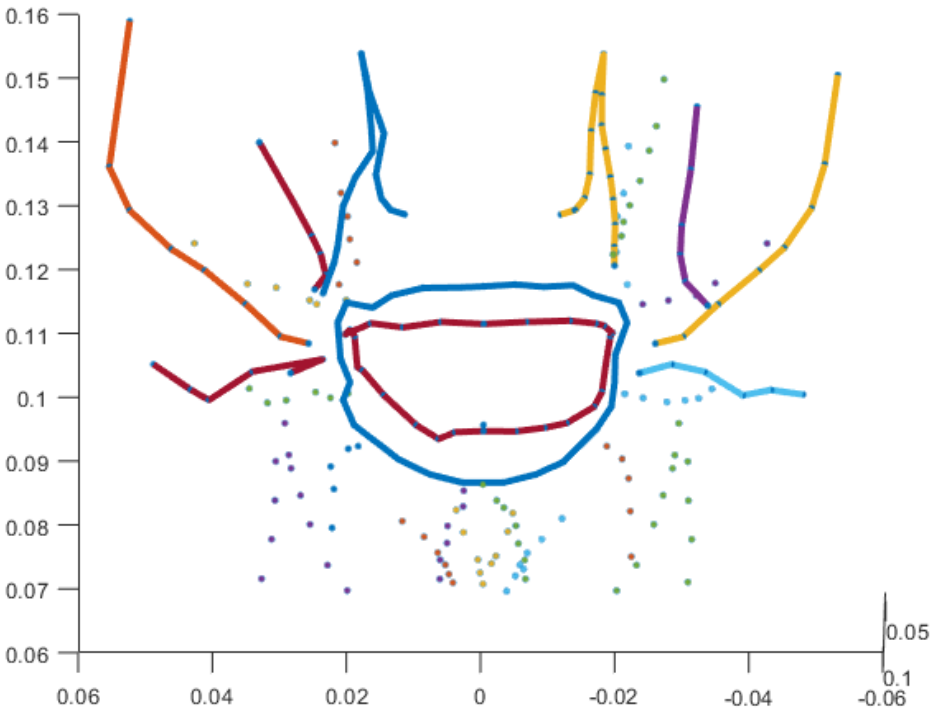


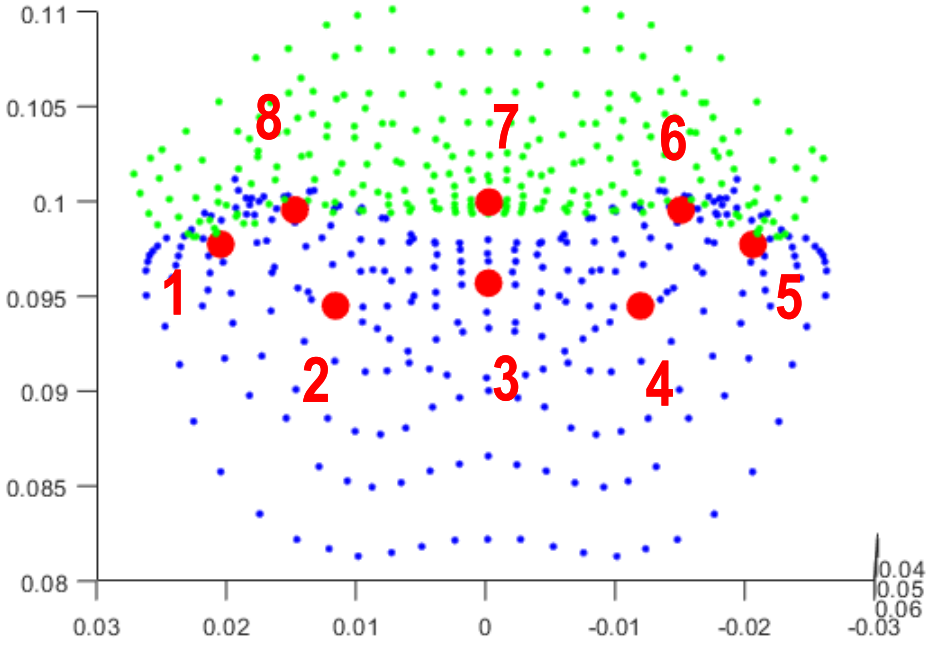
Figure 63. The network architecture of TD3: one actor network and two critic networks.

6.1.3.2. Reward Function, Action Space, and State Space

The aim of our study is to find the appropriate muscle excitations for performing a facial motion, which is generated by defining the appropriate biomechanics-inspired reward function. In our model, action is a vector of 10 pairs of left and right muscles in terms of muscle forces normalized between 0 and 1. To avoid the exhausting search, only significant muscles (*left and right Levator Anguli Oris (LAO), left and right Levator Labii Superioris Alaeque Nasi (LLSAN), left and right Zygomatics (ZYG), left and right Risorious (RIS), Orbicularis Oris Marginalis (OOM), Orbicularis Oris Peripheralis (OOP)*) were included in into training process (Figure 64a). In our model, the agent's state was defined through a set of landmark points focusing on the mouth region of the face. In fact, 8 key points on the lips are chosen as representations of the state of the face (Figure 64b).



(a)



(b)

Figure 64. Selected muscles excitations for training (a) and landmark points for the RL agent's state (b).

Regarding the reward function, the agent receives a reward value from the environment at each time step. Note that the training efficiency of the reinforcement learning algorithm depends strongly on defining the reward function. In our study, the reward function is

designed by a motion-oriented (e.g. symmetry-targeted motion, smile expression, sound pronunciations) strategy. More precisely, different reward functions were formulated using the Euclidean distance and angle created from the defined 8 landmark points. Mathematically, reward functions are defined as follows:

$$R_{symmetry}^{distance} = -1000 * \left(r_1^d + r_2^d + r_3^d\right) \quad (1)$$

$$R_{symmetry}^{angle} = -(r_1^a + r_2^a + r_3^a) \quad (2)$$

$$R_{smile} = \Delta d \quad (3)$$

where r_i^d, r_i^a are the symmetry value-based distance and angle between the left side compared to the right side for each pair point (point 1-5, 2-4, 8-6). Δd is the total moving up of point #1 and point #5 defined in Figure 64b.

6.1.4. Information exchange protocol and implementation

Our study is based on two modeling platforms (i.e. Artisynth and PyTorch) coming from different fields. The exchange information between these platforms needs a novel communication protocol. Artisynth-RL has been proposed to open the exchange capacity with rigid multi-bodies dynamics simulation [391]. In the present study, Artisynth-RL was extended to exchange information between the PyTorch platform which is a Python-based training platform for RL models, and Artisynth for running face finite element simulation. Note that Artisynth-RL is a cross-platform-based JavaScript. To achieve this objective, different technologies (RESTful API as a plugin, Spark framework from java, and Request package from python) were used as shown in Figure 65. Regarding the communication protocol, the muscle excitations from the reinforcement learning module are posted into the Artisynth module, then a new state is obtained from the Artisynth module to the reinforcement learning module after each simulation step.

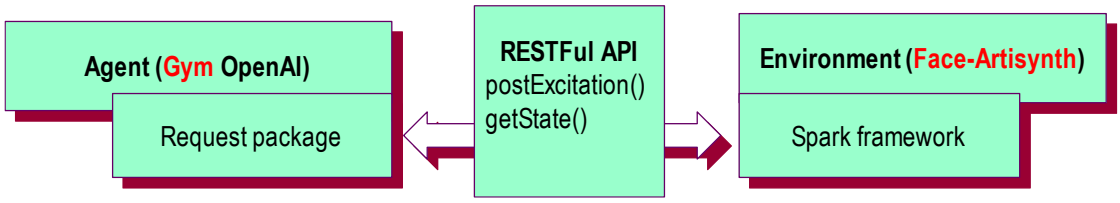


Figure 65. RESTful API as a plugin for bridging reinforcement learning and Artisynth

As hardware configuration, a virtual machine configuration with ubuntu 20.04, 8 CPU, 16 Gb RAM, Python 3.6, and the open-source stable-baselines3 was used for the training process.

6.1.5. Evaluation and validation

An open-access 3D face database, named Bosphorus, was used for evaluation and validation purposes. This database includes 105 subjects (44 females and 61 males) with different expressions, poses, and occlusion conditions. 65 subjects have 7 expressions such as happiness (smile), surprise, fear, sadness, anger, disgust, and neutral. Happiness (smile) and

neutral expressions, which are available in 130 face scans of all the subjects, were used for further validation. Firstly, three key points at positions #1, #5, and #7 (as in Figure 64b) were manually picked for each face scan. Secondly, all the face scans were transferred such that the point at position #7 is at the origin (coordinate [0, 0, 0]) and face scans of the same person are at the same orientation. Finally, the total displacement of moving up action of the two key points at positions #1 and #5 (as in Figure 64b) was computed by subtracting the corresponding points of the smile face scan and the neutral face scan of the same person. In fact, the values of the used reward functions were computed for each posture (neutral and smile expression). Obtained values were represented in mean and standard deviation and then compared to the final outcomes from the RL process.

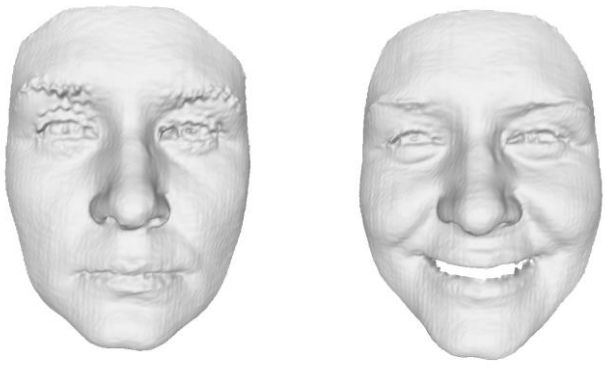


Figure 66. Illustration from Bosphorus database with two expressions: neutral (left) and smile (right).

A hyperparameter tuning process was implemented to select the best neural network architecture and parameters. In particular, the hyperparameter tuning process is not automatically tuned, but manually selects each parameter, while other parameters remain unchanged. Each trial was performed for a training task with the hyperparameter within a predefined set to ensure that the agent successfully explores and learns to make decisions in its environment. The predefined set of hyperparameters includes several most critical parameters, which govern the performance of reinforcement learning such as the neural network size (nodes in hidden layers ([64, 64] or [400, 300] for the actor and the critic networks)), learning rate (0.001, 0.01, note that the learning rate is shared for all networks), batch size (16, 32), τ parameter, which used to soft update both critic and actor target networks (0.001, 0.005).

6.2. Computational results

6.2.1. RL accuracy and performance

The reward and the loss evolutions during the training process of the agent to conduct one symmetry action were shown in Figure 67. In general, after more than 100 episodes of random interaction in the environment, the agent starts to learn from previous trials and can find the optimal policy after more than 300 episodes of training. In particular, the learning start parameter was set to 100 in the training phase, allowing the reinforcement learning agent to collect a set of transitions ($\mathcal{D} = (s, a, r, s', d)$) by performing a random action from the action space to the environment before learning from previous trials. These random actions result in an unstable trend in the reward values in the first 100 episodes. Having just 200 episodes of learning, the agent still learns and explores the environment to discover the optimal policy. During learning, the not optimal policy may predict the random actions for exploring more the environment that might dramatically drop reward values. From 200 to 300 episodes of learning, the agent gradually finds the optimal policy after more than 300 episodes of training. The reduction in actor loss and critic loss values demonstrates the efficacy of the learning strategy in both DDPG and TD3 methods.

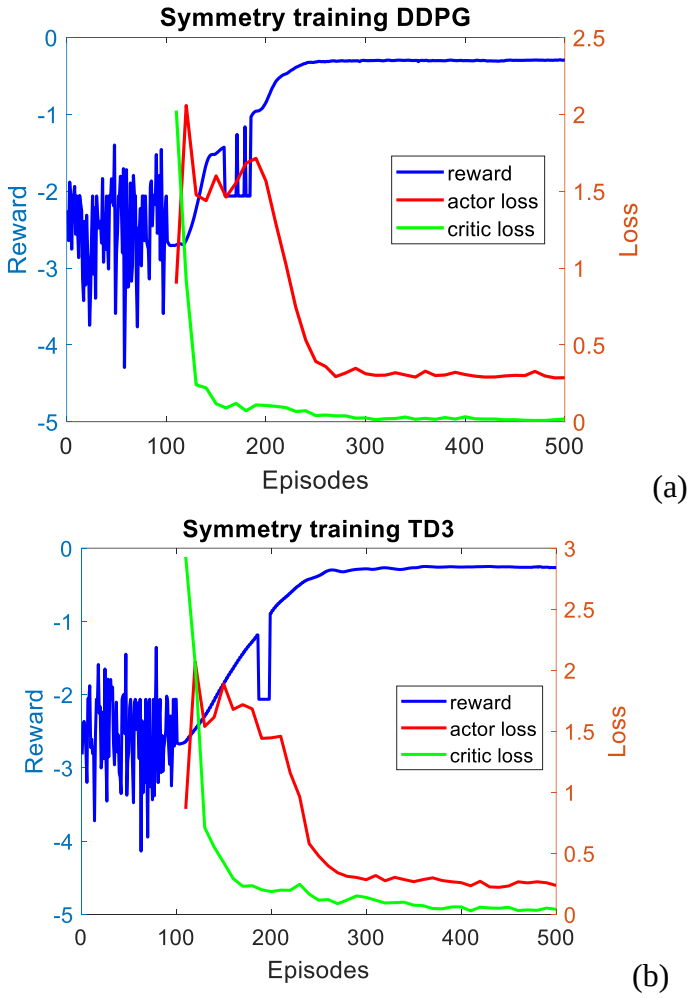


Figure 67. The reward value and the loss values of the actor network and the critic network during training reinforcement learning agent with two different methods: DDPG (a), TD3 (b) for symmetry-oriented functional rehabilitation using 4 muscles as ZYG, RIS, OOM, OOP.

The reward and the loss evolutions during the training process of the agent to perform the smile action were shown in Figure 68. Similar patterns are observed according to Figure 67. More precisely, the agent spends 100 episodes collecting a set of transitions by taking random actions to the environment resulting in instability of reward values. It then starts to learn from previous trials and can find the optimal policy after more than 200 episodes of training. The training is successfully demonstrated by the reduction of the loss value of both actor and critic networks.

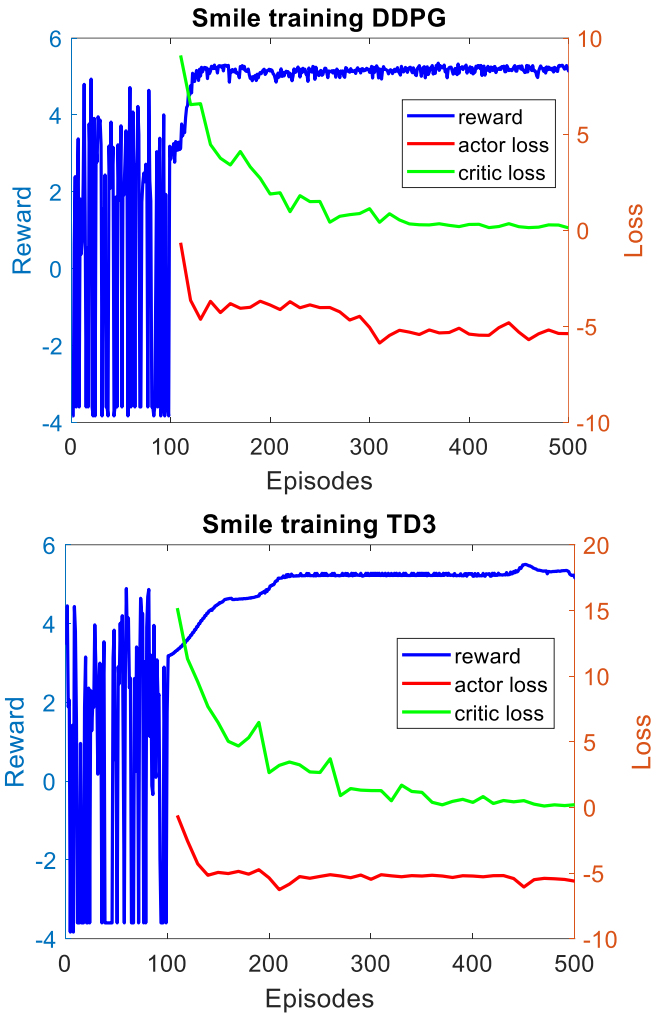


Figure 68. The reward value and the loss values of the actor network and the critic network during training reinforcement learning agent with two different methods: DDPG (left), TD3 (right) for smile-oriented functional rehabilitation using 3 facial muscles as LAO, LLSAN, ZYG.

The effect of the network architecture and τ parameter was reported in Figure 69. The network architecture has a more important effect than that of the τ parameter.

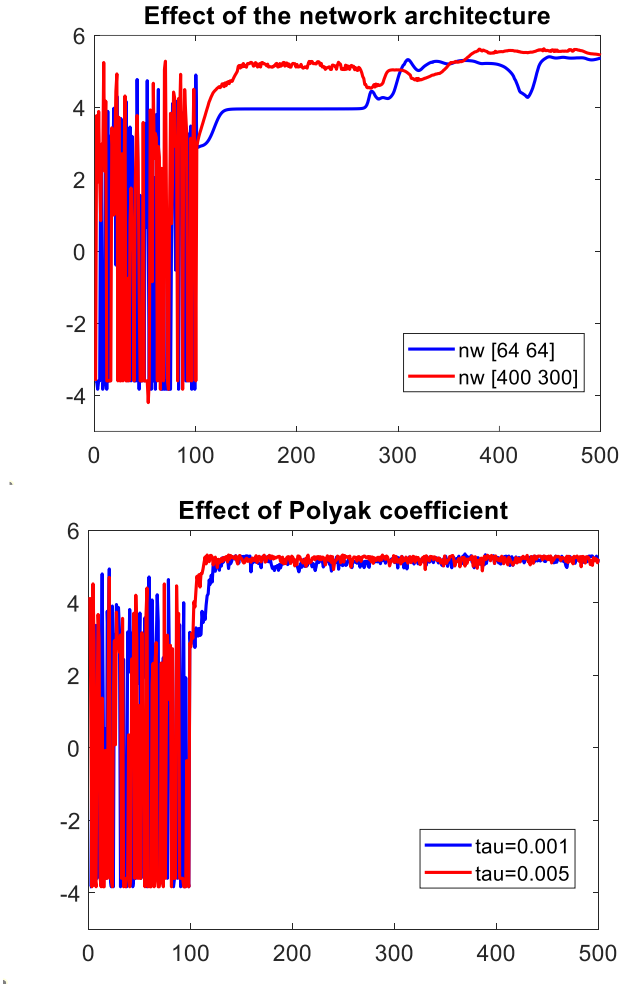


Figure 69. Reward value during training reinforcement learning agent with different hyperparameters and the network architecture.

The reward values were predicted by the hyperparameter tuning process for the DDPG method when training smile expression as shown in Table 11. The process can help to identify a better network architecture with an associated optimal set of hyperparameters. Based on the reward value, reinforcement learning with the architecture of two hidden layers [400, 300] for the actor and the critic networks, batch size 16, learning rate for all networks 0.001, τ parameter 0.005, and without action noise yields the best reward value. The computational time for training 500 episodes for each smile training and symmetry training is around 6 hours. However, the most computational time is in the simulation environment, where each simulation lasts for 30 seconds (5 hours for 500 episodes related to simulate and restart of Artisynth only).

Table 11. The reward values obtained during the hyperparameter tuning process

Hyperparameters		Reward
Batch size	16	5.21
	32	5.16
Learning rate	0.01	4.59
	0.001	5.35
Network architecture	[64, 64]	4.81
	[400, 300]	5.45
τ parameter	0.001	5.22
	0.005	5.26
Action noise	With	4.69
	Without	5.35

6.2.2. Facial motion learning

The obtained outcome and corresponding muscle excitation value of the symmetry-oriented functional rehabilitation are shown in Figure 70. and *Table 12*. According to the prediction, the muscle on the right side of the mouth such as right OOM, right RIS, and right ZYG are activated, while only the left OOP muscle is activated to improve the symmetry of the face from the initial state with the reward value R = -2.06 to the new state with the reward value R = -0.23, which counts 88.8% improvement.

Figure 70. Face animation for symmetry-oriented motion. The face at the initial state (on the left R = -2.06) and after receiving muscle excitation (on the right R = -0.23) output from reinforcement learning for symmetry-oriented functional rehabilitation.

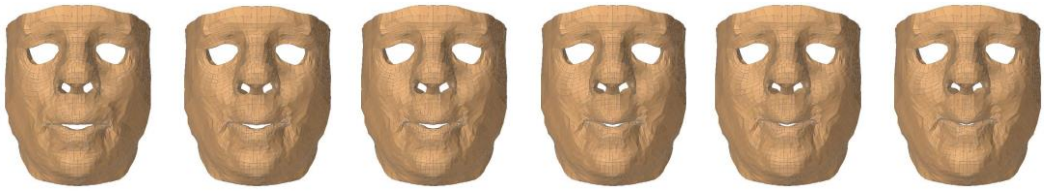


Figure 71. Face animation for smile-oriented motion. The face at the initial state (on the left R = -1.6) and after receiving muscle excitation (on the right R = 5.35) output from reinforcement learning for smile-oriented motion.

Regarding the smile-oriented motion simulation of the finite element model of the face, both left and right muscles of LAO and ZYG are activated, while LLSAN is not activated as in Figure 71 and *Table 13*. The measured reward value increases from -1.6 at the initial state to 5.3 at the terminal state. The obtained muscle activation levels for smiling movement are within the range of values reported by Flynn et al. [122]. However, it is important to note that there is a difference in smiling patterns between our simulation (i.e. unconstrained smile) and their simulations (i.e. miles with an open mouth or closed mouth).

The muscle action line length change and contraction amplitude $\xi^{CE} = \frac{L-L_0}{L_0} = \frac{\Delta L}{L_0}$ are shown in *Table 12*. The contraction amplitudes of OOM and OOP are estimated as the area that these muscles cover $\xi^{CE} = \frac{S-S_0}{S_0} = \frac{\Delta S}{S_0}$. Related to the smile, the right ZYG contracts -16.26%, while this number on the left is -15.12%. The right and left LAO contract around -30%. Note that all muscle contraction levels during smiling are in good agreement with those estimated using Kinect-driven rigid multi-bodies modeling (Nguyen et al. [109] (2021)).

Table 12. Muscle contraction levels during different facial expressions and comparison to the literature data.

Muscle / ξ^{CE}	symmetry		Smile		Nguyen et al. [109] (smile)
	$L_0\ (mm)$ / $S_0\ (mm^2)$	$\frac{\Delta L}{L_0}$ (%) / $\frac{\Delta S}{S_0}$ (%)	$L_0\ (mm)$ / $S_0\ (mm^2)$	$\frac{\Delta L}{L_0}$ (%) / $\frac{\Delta S}{S_0}$ (%)	$\frac{\Delta L}{L_0}$ (%)
Right ZYG	52 mm	-2.19	54.5 mm	-16.31	From -9.13 to -19.72
Left ZYG	52.2 mm	1.12	54.6 mm	-15.09	From -13.59 to -21.32
Right LLSAN	27.2 mm	-1.62	27.9 mm	-10.26	From -1.99 to -8.12
Left LLSAN	27.2 mm	-0.38	27.9 mm	-10.2	From -0.69 to -6.13
Right LAO	27.3 mm	-1.18	30 mm	-28.13	From -18.66 to -29.46
Left LAO	24.3 mm	1.56	30 mm	-28.17	From -21.19 to -28.03
Right RIS	52.2 mm	-6.21	52.9 mm	-8.12	From 3.55 to 7.30
Left RIS	52 mm	2.44	52.9 mm	-8.135	From -3.09 to 6.96
OOM	590 mm^2	-12.97	665 mm^2	-2.01	-
OOP	1099 mm^2	-7.98	1138 mm^2	17.40	-

Table 13. Muscle activation levels reported from our simulation and its comparison to the literature data

Muscle	symmetry	smile	Flynn et al. [122] (closed mouth smile)	Flynn et al. [122] (open mouth smile)
Right ZYG	0.2	0.4	0.2	0.5
Left ZYG	0	0.4	0.2	0.5
Right LLSAN	0	0	0.1	0.5
Left LLSAN	0	0	0.1	0.5
Right LAO	0	0.4	0.1	0.5
Left LAO	0	0.4	0.1	0.5
Right RIS	0.4	0	0.2	0.6
Left RIS	0	0	0.2	0.6
Right OOM	0	0	0	0
Left OOM	0	0	0	0
Right OOP	0.1	0	0	0
Left OOP	0.4	0	0	0

6.2.3. Evaluation and validation

For symmetry-oriented motion, the muscle excitations predicted by the trained agent help to increase the value of reward from R = -2.06 to R = -0.23, which counts for ~88.8%. While this number for smile-oriented motion, the reward value increases from R = -1.6 at the initial state to R = 5.3 at the terminal state, which is 0.35 cm moving up on average for each corner of the mouth. This is within the range of movements compared to the value calculated from the Bosphorus database, this value is $0.4 \pm 0.32\ cm$ when a person makes maximum effort to smile as in Figure 72.

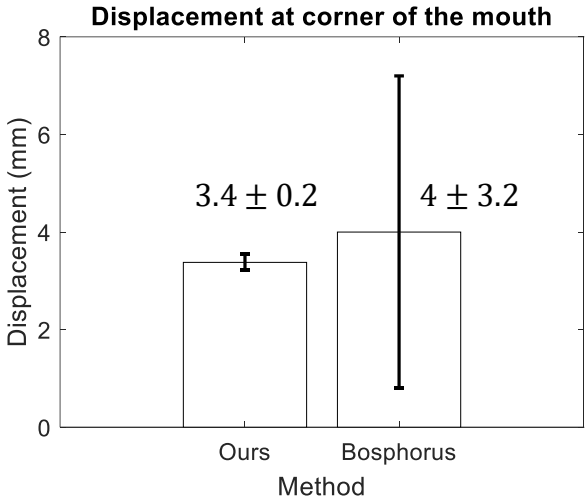


Figure 72. Displacement of the corner point of the mouth (moving up direction) of our method and from the Bosphorus database of the smile position compared to the neutral position.

6.3. Discussion

Understanding the muscle coordination mechanism of facial expressions plays a crucial role in the facial rehabilitation interventions for facial palsy or facial transplantation patients. Numerical models (i.e. finite element models) have been intensively developed [15] to provide a better understanding of this complex process. However, these developed models are descriptive and their predictive capacity is still limited. Besides, computer-based monitoring systems, that automatically recognize action units (AUs) to provide quantitative and objective information on the facial motion during the rehabilitation exercise have been developed [40], [365], [392]. Despite many efforts, understanding facial motion mechanism still remains a scientific and clinical challenge to help the involved patients to recover functional facial movements. In particular, the role of muscle excitation and its value for performing a desired facial movement for facial rehabilitation is still an open and longstanding research question. To achieve this complex and challenging objective, the present study aimed to couple reinforcement learning to a muscle-driven biomechanical model of the face to explore the facial motion learning capacity such as symmetry and facial smiling actions. For the first time, facial expressions (e.g. smile) are simulated without a priori input data (e.g. motion capture data). In fact, our novel coupling scheme allows to explore emerging properties of the facial muscle contraction mechanism and guides iteratively the face to express smiling action or becomes a symmetric face. Thus, obtained outcomes showed the potential application of this novel approach with facial palsy patients for a better understanding of facial muscle coordination and muscle activation patterns to target a specific motion.

More precisely, regarding the symmetry motion of the face, the muscle contraction involves right OOM, right RIS, and right ZYG, while only the left OOP muscle is activated. This is reasonable due to the physical-based model of the face used for symmetry training is drooping of the mouth on the right side. From the biomechanics point of view, this is a symptom of the facial palsy patient on the affected side of the face. In the smile-oriented motion of the face, levator anguli oris and zygomaticus are the main muscles responsible for smile action resulting in two corner points in the mouth moving up 0.35 cm, which is within the range of motion compared to the Bosphorus database (0.4 ± 0.32 cm). There is also a good agreement in muscle involvement for smile training as LAO and ZYG compared with the simulation of Flynn et al. [122] (2015). Note that Flynn et al. manually adjusted muscle excitation value to find the appropriate value for expression movements of the finite element model of the face. In fact, our present study revealed the usefulness of mechanical modeling coupled with reinforcement learning to guide the design of patient-specific precision rehabilitation for the face with muscle activation and coordination mechanisms.

Recently, deep reinforcement learning becomes an interesting solution for complex control problems [348]. The coupling between a reinforcement learning strategy and a deep neural network allows the agent to build knowledge by gathering information while interacting with the environment. In fact, no prior data is required for training. This particular character enhances the predictive capacity of the involved model. One of the challenges when developing an efficient RL model relates to the use of cumulative rewards to quantify how

the agents ought to take actions in an environment. In our present study, specific rewards were defined to guide the motion patterns toward the specific targets (symmetry and smiling motions). Moreover, two state-of-the-art RL methods (DDPG method and the successor TD3), which are off-policy algorithms and applicable for complex environments with continuous action spaces, were used to drive the face toward the targeted motions from the activation of the facial muscles. Note that the use of these methods leads to the win of the Learn to Move competition [393].

One of the most important limitations of the present work deals with the sensitive nature of the hyperparameters of the physics-based face model. Further investigations should be done to take the uncertainties of these parameters into account to provide more reliable prediction outcomes. In fact, each parameter should be represented in a more generic format like interval or probability-based structures (e.g. probability density function (PDF); cumulative distribution function (CDF)), and then associated outcomes (i.e. muscle activation) should be estimated within a plausible range of values. However, taking the parameter uncertainty into account increases drastically the computational cost during the reinforcement learning process. Thus, more efficient uncertainty propagation algorithms should be investigated to cope with this constraint. Moreover, the present face model includes only 10 muscles. In particular, all parameters were set up for a generic model. Thus, a more detailed face model and patient-specific properties of the facial tissues and structures should be taken into consideration from medical imaging toward a patient-specific rehabilitation application. In particular, the increase in the number of muscles of interest will allow the modeling system to explore full muscle action patterns of the face. Thus, the present system could benefit from the FACS pattern for a given expression to converge quickly to the optimal solution and then other applications like speech synthesis or language learning could be investigated. Regarding the limitation of the used RL approach, the use of only a deep neural network seems to be underestimated for the complex face motion coordination. As perspective, a multi-network approach should be investigated for a better coordination of the facial muscle activations and contractions. Finally, the coupling between RL and FE modeling frameworks requires the development of a specific communication protocol. In further work, the developed information exchange protocol will be improved to provide a generic communication channel between the RL framework and any other powerful and dedicated FE modeling frameworks like Abaqus or Ansys to overcome the limitation of the current physics-based face model.

6.4. Conclusions

Facial palsy patients or patients with facial transplantation have abnormal facial motion due to altered facial muscle functions and nerve damage. Computer-aided systems and physics-based models have been developed to provide objective and quantitative information. However, the predictive capacity of these solutions is still limited to explore the facial motion patterns with emerging properties. The present chapter explored the muscle excitation patterns by coupling reinforcement learning with a finite element model of the face. We developed, for the first time**, a novel coupling scheme to integrate the finite element simulation into reinforcement learning for facial motion learning**. In particular, two state-of-the-art reinforcement learning algorithms (deep deterministic policy gradient (DDPG) and Twin-delayed DDPG (TD3)) were successfully applied and implemented to drive the simulations of symmetry-oriented and smile movements. Obtained results were in very good agreement with the experimental observation. In fact, a better understanding of the facial muscle activation and coordination mechanism is of great clinical interest to guide the optimal rehabilitation strategy. The present work opens new avenues to achieve this challenging objective.

A paper for this chapter “Reinforcement learning coupled with finite element modeling for facial motion learning” is currently under the first revision in the Computer Method and Program in Biomedicine journal (Q1, IF@2022 = 5.428).

As perspectives, this method will be applied for subject-specific patient models derived from numerical models derived from medical images performed in our team [15], [16], [338]. The present chapter will finally be applied to build a decision support system for facial palsy and facial transplantation patients to guide and optimize the functional rehabilitation program.

Chapter 7

General Discussion

This chapter will provide a general discussion, main contributions as well as current limitations of the thesis.

7.1. Thesis overview

Facial expression contributes a big part to human communication, which is essential to human development and personal life [1], [7]. Furthermore, facial expression also has an important role in a person's identity as well as health [9]. In fact, facial palsy patients or patients with facial transplantation have abnormal facial motion patterns due to altered facial muscle functions and nerve damage. This leads to abnormal motion control for different movements such as eating, speaking, or facial expressions as well as the asymmetric face [10]–[13]. As a result, these involved patients have had their personal, professional, and social lives negatively impacted [10]–[13]. Thus, the restoration of normal and symmetrical facial expressions would considerably improve the quality of life and social interactions for involved patients. An appropriate facial rehabilitation program is a critical clinical stage that influences the efficacy of surgical and pharmacological treatments [20], [25], [27], [394]–[396]. Especially, involved patients should have their facial paralysis levels assessed before being referred to appropriate rehabilitation programs. It is important to note that traditional facial rehabilitation has mainly based on a mirror approach to monitor the visual qualitative feedback from the rehabilitation exercise. More precisely, patients watch their distorted features in the mirror as a reference to teach themselves the right expressions during rehabilitation exercises. This strategy is ineffective and subjective without any additional feedback. Computer-aided systems based on physics-based models have also been developed to provide objective and quantitative information. However, there are still developments that could be done to overcome the limitations of existing clinical support systems:

1) The lack of building 3D information from images or depending strongly on the selected cameras.

2) The lack of analyzing the face in terms of expression recognition and symmetry analysis.

3) The lack of making use of artificial intelligence (reinforcement learning, for example) for facial rehabilitation guidelines using biomechanical knowledge. This is the limitation of the predictive capacity of the facial motion patterns with emerging properties.

Consequently, the objective of this Ph.D. is to develop innovative engineering solutions toward a next-generation computer-aided decision support system for facial analysis and rehabilitation.

7.2. Main contributions

The thesis provided four main contributions:

The first contribution concerns the fast reconstruction of the 3D face shape from a single 2D image. *This supports to provide reliable feedback for clinical decision support, which is portable, easy to use, cheap, and independent of the choice of devices*.

The second and third contributions concern the improvement of facial expression recognition and facial symmetry based on 3D point sets. *These works could potentially help clinicians determine the severity of facial paralysis as well as the level of recovery of the rehabilitation process.*

The final contribution is the proposition of a novel modeling workflow for learning facial motion by coupling reinforcement learning and finite element modeling for facial motion learning and prediction. *This part is of help for guiding patients on how to conduct rehabilitation exercises by providing visualization of correct rehabilitation exercises*.

7.2.1. 3D face reconstruction from a single image for facial palsy patients

The 3D reconstruction of an accurate face model is essential to provide reliable feedback for clinical decision support for facial disorders. Thus, this allows analyzing the face with external (i.e. face deformation) and internal (i.e. facial muscle mechanics) feedback for the diagnosis and rehabilitation process of facial palsy and facial transplantation patients [33]. The facial analysis for diagnosis and treatment has mainly been based on 2D images [365]–[367]. It remains a challenge due to the variations in pose, expressions, and lighting conditions. 3D facial data can be acquired from medical imaging [121], [122], 3D scanners [112], [113], stereo-vision systems [123], or RGB-D devices as Kinect. The use of medical imaging leads to a very accurate 3D model but this is not appropriate for an easy-to-use, cheap and portable system. The use of depth cameras like Kinect can lead to a reasonable accuracy level while keeping the cheap cost, easy-to-use and portable requirements but the developed system depends strongly on the selected sensors. In fact, this could alter the future applicability due to stopped production like in the case of the Kinect V2 camera. Thus, it is necessary to have a more flexible and open method to build 3D information rather than using specific scanning devices. The recent development of deep learning (DL) models opens new challenges for 3D shape reconstruction from a single image. In fact, three distinguish approaches have been applied for reconstructing 3D shapes from 2D information. The first approach uses the statistical model fitting with a prior 3D facial model to fit the input images [129]–[131]. The second approach is based on the photometric stereo. The method is suitable for multiple images. The method combines a 3D template face model with photometric stereo methods to compute the surface normal of the face [143], [144]. The third approach uses deep learning to learn the shape and appearance of the face by training 2D-3D mapping functions [151], [152]. These approaches lead to very good accuracy levels for 3D face reconstruction of the subject-specific face. However, the 3D face shape reconstruction of facial palsy patients is still a challenge and this has not been investigated.

The objective of Chapter 3 was to apply these state-of-the-art methods to reconstruct the 3D face shape models of the facial palsy patients in natural and mimic postures from one single image. Three different methods (3D Basel Morphable model and two 3D Deep Pre-trained models) were applied to the dataset of two healthy subjects and two facial palsy patients. The methodology was also applied to reconstruct the 3D face of patients from 2D images collected from the internet. The first used method fits a 3DMM to a single image based on a scale orthographic projection [129]. The second used method reconstructs the 3D face based on an established FLAME head model [357]. The third applied method reconstructs the 3D face of the patient with weakly-supervised learning to regress the shape and texture coefficients from a given input image [361]. Reconstructed outcomes were compared to the 3D shapes reconstructed using Kinect-driven and MRI-based information. As result, the best mean error of the reconstructed face according to the Kinect-driven reconstructed shape is $1.5 \pm 1.1\ mm$. The best error range is $1.9 \pm 1.4\ mm$ when compared to the MRI-based shapes. Obtained results showed a very good reconstruction accuracy level compared to the Kinect-driven and MRI-based outcomes. This present study opens new avenues for the fast reconstruction of the 3D face shapes of the facial palsy patients from a single 2D image.

7.2.2. A novel solution for facial expression recognition

Facial expression recognition plays an essential role in human conversation and human-computer interaction. Recognizing human expression assists humans in face-to-face communication for interpreting the other's intentions. Mehrabian (1975) illustrated that the nonverbal components (such as facial expression) part of a speaker can account for 55% of the interpretation in the conversations, while the verbal part (i.e. part that relates to words) and the vocal part (i.e. part that relates to the sound) contribute only to 7% and 38%, respectively [8]. There is also a wide range of applications of facial expression recognition in the human-computer interaction [115], especially for the virtual reality and augmented reality systems [124], [180], and healthcare systems (e.g. facial nerve grading [181]). Most existing facial expression recognition in the past several decades has been based on 2D processing approaches [263]–[265]. The recognition performance based on 2D images remains challenging when processing expressions with large variations in different poses and lighting conditions. This task is particularly challenging when performing feature engineering, which is time-consuming and subjective. In fact, 3D information such as the 3D point cloud data with an end-to-end deep learning algorithm should be used to deal with the recognition degradation. Some approaches have successfully explored 3D face for expression recognition tasks. Precisely, several studies project 3D data into the 2D image plane and handle expression recognition based on features extracted from the 2D image plane [285]–[288]. Some other approaches find a new presentation of the 3D face in terms of curvature descriptor [289], scale-invariant feature transform (SIFT) feature [290], and spatial distances and displacements between facial landmarks [291], [292]. This, however, transforms the data and often reduces the depth information on the face. To the best of our knowledge, there is no research study using directly the 3D point cloud for facial expression recognition.

In Chapter 4, facial expression was recognized by applying a new class of deep learning called geometric deep learning directly on 3D point cloud data. In particular, two databases (Bosphorus and SIAT-3DFE) were used. The Bosphorus database includes sixty-five subjects with seven basic expressions (i.e. anger, disgust, fearness, happiness, sadness, surprise, and neutral). The SIAT-3DFE database has 150 subjects and 4 basic facial expressions (neutral, happiness, sadness, and surprise). Firstly, pre-processing procedures such as face center cropping, data augmentation, and point cloud denoising were applied to 3D face scans. Secondly, a geometric deep learning model called PointNet++ was applied. A hyperparameter tuning process was performed to find the optimal model parameters. Finally, the developed model was evaluated using the recognition rate and confusion matrix. The facial expression recognition accuracy on the Bosphorus database was 69.01% for 7 expressions and could reach 85.85% when recognizing five specific expressions (anger, disgust, happiness, surprise, and neutral). The recognition rate was 78.70% with the SIAT-3DFE database.

7.2.3. A novel solution for facial symmetry analysis

Analyzing facial symmetry is important to determine the severity of facial paralysis of involved patients. In fact, facial symmetry indicates the balance and equality of facial structure in terms of shape, size, location, and arrangement of left and right components on the sagittal plane [304]. Asymmetry face, on the other hand, depicts the bilateral difference between two sides. Facial symmetry is an important aspect of determining attractiveness and beauty [305]. Furthermore, patients with facial palsy and those who have had facial transplants have facial asymmetry, which results in unnatural facial expressions [31], [107]. As a consequence, facial symmetry analysis is vital to support the correct assessment and diagnosis of patients with facial palsy and facial transplantation. It can be of helps doctors in designing an effective individual rehabilitation program [21], [33].

Chapter 5 provided an indicator based on novel descriptors for analyzing the symmetry of the face and the difference between two ethnicities such as Caucasians and Asians. Two different facial point cloud databases (Bosphorus and SIAT-3DFE) corresponding to two different ethnicities were used. The Bosphorus database consists of sixty-five subjects with seven different expressions (i.e. anger, disgust, fearness, happiness, sadness, surprise, and neutral). The SIAT-3DFE database contains 150 subjects with four facial expressions (neutral, happiness, sadness, and surprise). Firstly, pre-processing procedures such as splitting the face into six parts (left and right eye, left and right nose, left and right mouth), and space normalization were applied on 3D face scans. Secondly, a geometric deep learning model called PointNet++ was applied to extract the features from the 3D point cloud data. Finally, these features extracted from the PointNet++ model were transformed using principal component analysis (PCA). The method aims to reduce the dimension and still keep the significant information. The descriptors were then tested whether they can be used for analyzing the symmetry and asymmetry of the face and in the meantime compared the difference between Caucasians and Asians. These descriptors could potentially be used for analyzing the symmetry of the face of facial palsy patients.

7.2.4. A novel solution for facial motion learning based on learning muscle-driven motion

The recovery of a symmetric face with balanced functionalities requires a complex rehabilitation process in which patients must practice patient-specific facial movements. Thus, understanding of facial motion mechanism helps the involved patients to recover symmetrical movements and normal facial expressions. It is important to note that current facial rehabilitation has mainly been based on a mirror approach to monitor the visual qualitative feedback from the rehabilitation exercise. More precisely, patients watch their distorted features in the mirror as a reference to teach themselves the right expressions during rehabilitation exercises. This strategy is ineffective and subjective without any feedback. Moreover, the current rehabilitation process is limited by a lack of patient-specific knowledge about muscles driving facial motions. Therefore, understanding facial motion mechanisms, muscle activation, and coordination are clearly fundamental. In addition, for investigating muscles driving facial motion problems, biomechanical models are recommended because they can be customized to reflect the true anatomy and pathological anatomical deformations as well as imitate physical processes [331]. In fact, physics-based facial models using finite element methods have been intensively developed to explore the role of facial muscle excitation, contraction, and coordination during facial motion [15], [16], [122], [328], [332]–[337]. Recently, the reinforcement learning strategy has been coupled with rigid multi-bodies dynamics to explore the motion of the lower limbs during walking and age-related falls in our team [348]. Thus, this learning strategy opens new avenues to explore human system motion and novel emerging properties without any a priori motion data.

Chapter 6 aims to explore the facial motion learning capacity by the coupling between reinforcement learning and finite element modeling. The main objective is to provide, for the first time, the modeling workflow for this complex coupling and then to evaluate different learning strategies to establish motion patterns of the face during facial expression motions. A physically-based model of the face within the Artisynth modeling platform was used. Information exchange protocol was proposed to exchange capacity between reinforcement learning and rigid multi-bodies dynamics simulation. Two reinforcement learning algorithms (deep deterministic policy gradient (DDPG) and Twin-delayed DDPG (TD3)) were used and implemented to drive the simulations of symmetry-oriented and smile movements. Numerical outcomes were compared to experimental observations (Bosphorus database) for evaluation and validation purposes. As result, after more than 100 episodes of exploring the environment, the agent starts to learn from previous trials and can find the optimal policy after more than 300 episodes of training. Regarding the symmetry-oriented motion, the muscle excitations predicted by the trained agent help to increase the value of reward from R = -2.06 to R = -0.23, which counts for ~89% improvement of the symmetry value of the face. For smile-oriented motion, two points at the edge of the mouth move up 0.35 cm, which is within the range of movements estimated from the Bosphorus database (0.4 ± 0.32 cm). Our

novel solution will explore the patient-specific facial motions without a priori data from the patient and then provides a set of facial muscle activation and coordination patterns for a specific rehabilitation-oriented movement (e.g. symmetry or smile).

7.3. Current limitations

In this project, several limitations related to each part were discussed.

Regarding the 3D face reconstruction, one important limitation of the present study deals with a small number of subjects and patients used for prediction. **Another limitation deals with the lack of facial palsy patients in the learning database and this has to be built with the clinicians**. This results in reducing several facial palsy patient's features (e.g. asymmetric face, dropping mouth corner and cheek) while reconstructing their 3D face. Thus, a larger and more diverse 3D facial database including facial palsy subjects should be acquired to confirm our findings and toward a potential clinical application. Another limitation of the study relates to the usage of the 3D statistical facial model. The first method used a 3DMM was based on the PCA basis vectors so that the reconstruction of more detailed information such as expressions and wrinkles can become a hard task. The second and third methods improve that by building a more diverse model with subtler information such as expressions and wrinkles. But these methods still use a linear model that could generate more errors due to facial shape variations, which cannot be modeled perfectly using a combination of linear components as noted in [128], [369], [370]. Improving the existing 3D face model can be a potential suggestion for future works.

In facial expression recognition and facial symmetry analysis parts, despite PointNet++ can work directly on 3D point clouds, 3D raw point clouds still require several preprocessing tasks such as removing the noise and cropping the face in the center region and data normalized. Cropping point cloud into the center region of the face reinforces the discriminative capability of expression feature for the deep learning model. In addition, the augmentation technique is also applied to enhance the size of our database. The main limitation of this part relates to the size of the Bosphorus and SIAT databases with high-quality data. Only 65 subjects with a total of 455 face scans (Bosphorus database) and 70 subjects with 280 face scans (SIAT database) were used for facial expression recognition, which can narrow the generalization of the method. Therefore, to be potentially applied for clinical applications, a larger 3D facial database including the facial palsy patient's database is required for confirming and improving the obtained result. Moreover, the level of expression and symmetry have not been estimated. It could help clinicians to determine the severity of facial paralysis as well as the level of recovery of the rehabilitation process.

In the facial motion learning part, one of the most important limitations of the present thesis deals with the sensitive nature of the hyperparameters (such as the number of muscles, the boundary of muscle excitation, and skin parameters) of the physics-based face model. Further investigations should be done to take the uncertainties of these parameters into account to provide more reliable prediction outcomes. Furthermore, subject-specific patient models should be implemented derived from numerical models derived from medical images performed by our team [15], [16], [338]. In fact, patient-specific properties of the facial tissues and structures were considered and are of importance for patient-specific rehabilitation applications. The present face model includes only 10 muscles. In particular, all parameters were set up for a finite element model of a generic face model. The increase in the

number of muscles of interest will allow the modeling system to explore full muscle action patterns of the face.

Chapter 8

Conclusions and Perspectives

This chapter presents conclusions as well as perspectives that need to be done for developing a next generation computer-aided decision support system for facial palsy and facial transplantation for optimizing the rehabilitation process.

8.1. Conclusions

In this thesis, several innovative engineering solutions were investigated toward a next-generation computer-aided decision support system for facial mimic analysis and rehabilitation. Firstly, the purpose of the thesis is to provide reliable feedback for clinical decision support by reconstructing the 3D face of the patient from 2D images. Secondly, the thesis could help clinicians to determine the severity of facial paralysis as well as the level of recovery of the rehabilitation process with facial expression recognition and facial symmetry analysis. Finally, the thesis could potentially guide patients on how to conduct rehabilitation exercises using the facial motion learning part.

Four main contributions have been proposed for addressing these researches. In particular, the 3D face of the patient can be reconstructed from a single 2D image using three different methods (morphable model and two deep learning models). Then, the 3D face can be analyzed in terms of facial expression recognition and facial symmetry analysis based on 3D point cloud data that can support facial paralysis diagnosis. Finally, we explored the muscle excitation patterns by coupling reinforcement learning with a finite element model of the face that can guide the patient to practice and find the optimal rehabilitation strategy.

There were several limitations about this thesis. A small number of data as well as lacking facial palsy patient data in the learning database lead to narrow the generalization of the method. The research used an available physical model of the face, which is sensitive to hyperparameters. A facial database including facial palsy and facial transplantation patients could be collected in the future for improving the obtained results. A more detailed face model and associated patient-specific properties of facial tissues and structures should be taken into consideration from medical imaging toward a patient-specific rehabilitation application. In addition, more rehabilitation exercises will be investigated for enhancing the facial rehabilitation process. Finally, the present framework will be applied to facial palsy and facial transplantation patients to guide and optimize the functional rehabilitation program.

8.2. Perspectives

Future works will be conducted to improve the results as well as overcome the existing limitations.

Regarding the 3D face reconstruction from 2D images for facial palsy patients, a 3DMM model built from a database with diverse ethnicities, wide ranges of ages, numerous numbers of subjects, and including facial palsy patients should be developed. This model could better capture the facial palsy patient features such as the asymmetric property, dropping mouth corner and cheek. *Regarding 3D facial expression recognition and facial symmetry analysis*, a similar 3D face dataset to build 3DMM should be developed with more subjects, diverse ethnicities, wide ranges of ages, and including facial palsy patients. Combining 3D point

cloud data and the texture of 2D images is also a suggestion for improving the accuracy of the prediction. *Regarding the facial motion learning by coupling reinforcement learning and finite element model of the face*, the next step would be to use realistic physical-based models of patient-specific faces for reinforcement learning algorithms. In that case, personalized treatment and rehabilitation planning could be provided.

The work can also be extended **to be able to apply to clinical decision support systems for facial diagnosis and facial rehabilitation**. **For better diagnosis**, the level of facial expression and facial symmetry should be estimated. This could be of helps clinicians to determine the severity of facial paralysis as well as the level of recovery of the rehabilitation process. This could be done automatically by labeling facial expressions and facial symmetry at different levels and then applying machine learning and deep learning for multiclass classification problems. **For better guiding patients during facial rehabilitation exercises**, more rehabilitation movement exercises such as facial expression and sound pronunciation should be learned in muscle-driven facial motion learning problems. This could be done by designing more reward function-oriented rehabilitation exercises for conducting reinforcement learning algorithms. Different reinforcement learning algorithms and transfer learning algorithms could be applied for improving the predicting capacity of muscle-driven facial movement.

Moreover, a novel framework will be proposed namely REHAB_DEEPFACE (Figure 73), which describes **the enhanced facial behavior recognition and rehabilitation using biomechanical features and deep learning approaches**. The system is designed to deliver a set of external (i.e. face deformation, symmetry, expression) and internal (i.e. facial muscle mechanics) feedback for both doctors and patients. Firstly, a single 2D image of the patient will be captured by any device (e.g. camera, phone, …). Secondly, a 3D geometrical model of the face will be reconstructed. This face model then will be analyzed in terms of expression and symmetry. The analysis of facial symmetry would allow quantifying the level of facial symmetry or asymmetry. Facial expression recognition can quantify the different expressions for facial mimics such as for example smile or pronunciation of sounds “o”, “pou” … Thank to the quantitative and objective indicators, **the severity of facial palsy patient can be diagnosed with the help of clinicians, and rehabilitation exercises will be planned. These exercises will be used to design reward-oriented rehabilitation functions**. Reward functions and the 3D patient-specific model will be applied for facial motion learning. **Facial motion learning-oriented rehabilitation exercises will provide the best facial mimic to be achieved**.

This visualization of the facial mimic model will be aligned with the texture of the face to be more realistic. **During practice, the patient will mimic and try to reach what he/she sees in the realistic visualization**. The process will be repeated again and again. We assumed that these repetitive rehabilitation exercises can make a change in the patient’s brain. In fact, these steps of ‘human learning’ will assist the brain, facial muscles, and facial nerve systems to reroute the electrical signals from the brain to muscles that were interrupted. Moreover, the similarity index will also be measured between the face of the patient and the

realistic finite smile face to estimate the recovery level of the facial palsy patient. This might improve the rehabilitation process.

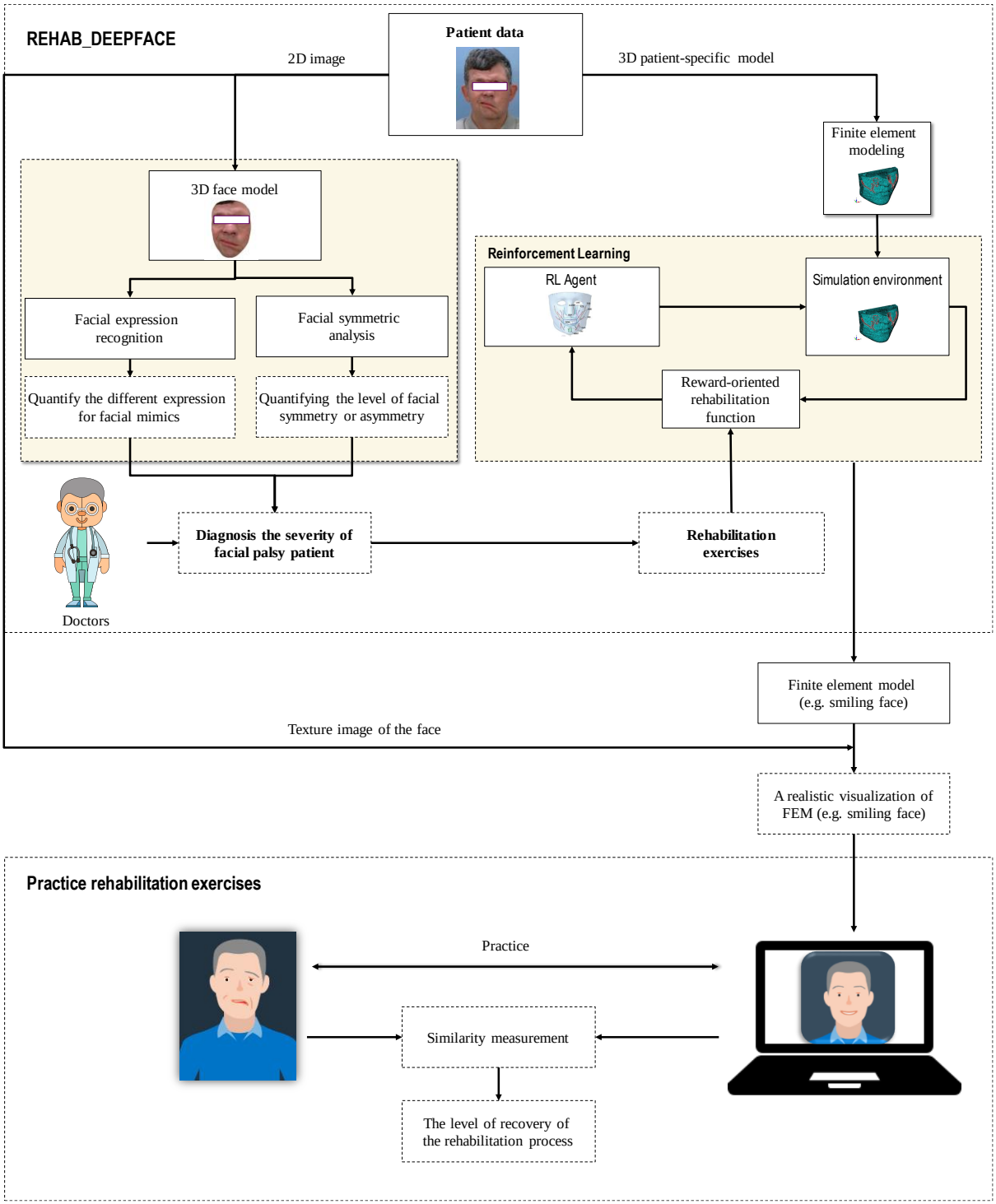


Figure 73. Framework of the decision support system for facial rehabilitation, REHAB_DEEPFACE (dash box: need to be done)

Some technological components of REHAB_DEEPFACE are analyzed for a potential transfer of technology [397]. In fact, the final product should be a software for a clinical decision support system for facial mimic rehabilitation to be used by patients and clinicians.

In that context, the system will improve the rehabilitation process and the quality of the life of the patients.

Publications

Journal articles:

Accepted

Duc-Phong Nguyen, Marie-Christine Ho Ba Tho, & Tien-Tuan Dao (2021). Enhanced facial expression recognition using 3D point sets and geometric deep learning. *Medical & Biological Engineering & Computing*, *59*(6), 1235-1244. (Q2, IF@2020=2.602). 2020 (2020) 1–30. DOI: https://doi.org/10.1007/s11517-021-02383-1.

Duc-Phong Nguyen, Marie-Christine Ho Ba Tho, & Tien-Tuan Dao, Reinforcement Learning coupled with Finite Element Modeling for Facial Motion Learning, *Computer Methods and Programs in Biomedicine* vol. 221, 106904 (Q1, SCI, IF=5.428)

Submitted

Duc-Phong Nguyen, Tan-Nhu Nguyen, Marie-Christine Ho Ba Tho & Tien-Tuan Dao. Fast 3D Face Reconstruction from a Single Image using Different Deep Learning Approaches for Facial Palsy Patients. *IRBM - Innovation and Research in BioMedical Engineering (Q2, SCIE, IF=1.856).*

In preparation

Duc-Phong Nguyen, Marie-Christine Ho Ba Tho & Tien-Tuan Dao. Facial Symmetry Analysis based on Novel Shape Descriptors between Two Different Populations. *Medical & Biological Engineering & Computing* (Q2, IF@2020 = 2.602).

Conference papers:

Accepted

Duc-Phong Nguyen, Marie-Christine Ho Ba Tho, & Tien-Tuan Dao, Reinforcement Learning coupled with Finite Element Modeling for Facial Motion Learning. 9th World Congress of Biomechanics (WCB2022), July 10–14, 2022, Taipei, Taiwan (Oral presentation).

Duc-Phong Nguyen, Marie-Christine Ho Ba Tho, & Tien-Tuan Dao, A framework of decision support system for facial rehabilitation based on Reinforcement Learning coupled with Finite Element Model. Virtual Physiological Human (VPH2022). September 6–9, 2022, Porto, Portugal (Oral presentation).

Report:

Duc-Phong Nguyen, Tien-Tuan Dao, and Marie-Christine Ho Ba Tho, "Valorization of REHAB_DEEPFACE," Report, Université de Technologie de Compiègne.

References

[1] C. Cherry, *On human communication: a review, a survey, and a criticism*, 3. ed. Cambridge, Mass.: MIT Pr, 1982.

[2] Judy Pearson, Paul Nelson, Scott Titsworth, and Angela Hosek, *Human Communication*.

[3] A. Mehrabian and M. Wiener, "Decoding of inconsistent communications," *J. Pers. Soc. Psychol.*, vol. 6, no. 1, pp. 109–114, May 1967, doi: 10.1037/h0024532.

[4] A. Mehrabian and S. R. Ferris, "Inference of attitudes from nonverbal communication in two channels," *J. Consult. Psychol.*, vol. 31, no. 3, pp. 248–252, Jun. 1967, doi: 10.1037/h0024648.

[5] A. Mehrabian, "Communication Without Words," in *communication theory*, 2nd ed., C. D. Mortensen, Ed. Routledge, 2017, pp. 193–200. doi: 10.4324/9781315080918-15.

[6] K. Kaulard, D. W. Cunningham, H. H. Bülthoff, and C. Wallraven, "The MPI Facial Expression Database — A Validated Database of Emotional and Conversational Facial Expressions," *PLoS ONE*, vol. 7, no. 3, p. e32321, Mar. 2012, doi: 10.1371/journal.pone.0032321.

[7] B. Ko, "A Brief Review of Facial Emotion Recognition Based on Visual Information," *Sensors*, vol. 18, no. 2, p. 401, Jan. 2018, doi: 10.3390/s18020401.

[8] A. Mehrabian, *Silent messages*, Nachdr. Belmont, Calif.: Wadsworth, 1975.

[9] C. Frith, "Role of facial expressions in social interactions," *Philos. Trans. R. Soc. B Biol. Sci.*, vol. 364, no. 1535, pp. 3453–3458, Dec. 2009, doi: 10.1098/rstb.2009.0142.

[10] L. E. Ishii, J. C. Nellis, K. D. Boahene, P. Byrne, and M. Ishii, "The Importance and Psychology of Facial Expression," *Otolaryngol. Clin. North Am.*, vol. 51, no. 6, pp. 1011–1017, 2018, doi: 10.1016/j.otc.2018.07.001.

[11] K. R. Bogart, L. Tickle-Degnen, and N. Ambady, "Communicating Without the Face: Holistic Perception of Emotions of People with Facial Paralysis," *Basic Appl. Soc. Psychol.*, vol. 36, no. 4, pp. 309–320, Jul. 2014, doi: 10.1080/01973533.2014.917973.

[12] G. Fuller and C. Morgan, "Bell's palsy syndrome: mimics and chameleons," *Pract. Neurol.*, vol. 16, no. 6, pp. 439–444, Dec. 2016, doi: 10.1136/practneurol-2016-001383.

[13] D. S. Grewal and D. S. Grewal, *Atlas of surgery of the facial nerve: an otolaryngologist's perspective.* New Delhi: Jaypee Brothers, 2012. Accessed: Feb. 11, 2022. [Online]. Available: http://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=706773

[14] T. Wu, A. P. L. Hung, P. Hunter, and K. Mithraratne, "Modelling facial expressions: A framework for simulating nonlinear soft tissue deformations using embedded 3D muscles," *Finite Elem. Anal. Des.*, vol. 76, pp. 63–70, 2013, doi: 10.1016/j.finel.2013.08.002.

[15] A.-X. Fan, S. Dakpé, T. T. Dao, P. Pouletaut, M. Rachik, and M. C. Ho Ba Tho, "MRI-based finite element modeling of facial mimics: a case study on the paired zygomaticus major muscles," *Comput. Methods Biomech. Biomed. Engin.*, vol. 20, no. 9, pp. 919–928, Jul. 2017, doi: 10.1080/10255842.2017.1305363.

[16] T. T. Dao, A. X. Fan, S. Dakpé, P. Pouletaut, M. Rachik, and M. C. Ho Ba Tho, "Image-based skeletal muscle coordination: case study on a subject specific facial mimic simulation," *J. Mech. Med. Biol.*, vol. 18, no. 2, pp. 1–15, 2018, doi: 10.1142/S0219519418500203.

[17] G. K. Scadding, P. D. Bull, and J. M. Graham, Eds., “The Facial Nerve,” in *Pediatric ENT*, Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, pp. 479–484. doi: 10.1007/978-3-540-33039-4_47.
[18] M. J. Warner, J. Hutchison, and M. Varacallo, “Bell Palsy,” in *StatPearls*, Treasure Island (FL): StatPearls Publishing, 2022. Accessed: Feb. 11, 2022. [Online]. Available: http://www.ncbi.nlm.nih.gov/books/NBK482290/
[19] M. Wernick Robinson, J. Baiungo, M. Hohman, and T. Hadlock, “Facial rehabilitation,” *Oper. Tech. Otolaryngol. - Head Neck Surg.*, vol. 23, no. 4, pp. 288–296, 2012, doi: 10.1016/j.otot.2012.10.002.
[20] J. M. Dubernard *et al.*, “Outcomes 18 months after the first human partial face transplantation,” *N. Engl. J. Med.*, vol. 357, no. 24, pp. 2451–2460, 2007, doi: 10.1056/NEJMoa072828.
[21] M. Lorch and S. J. Teach, “Facial Nerve Palsy: Etiology and Approach to Diagnosis and Treatment,” *Pediatr. Emerg. Care*, vol. 26, no. 10, pp. 763–769, Oct. 2010, doi: 10.1097/PEC.0b013e3181f3bd4a.
[22] M. R. Garanhani, J. Rosa Cardoso, J. R. Cardoso, A. de M. G. Capelli, and M. C. Ribeiro, “Physical therapy in peripheral facial paralysis: retrospective study,” *Braz. J. Otorhinolaryngol.*, vol. 73, no. 1, pp. 106–109, Feb. 2007, doi: 10.1016/s1808-8694(15)31131-9.
[23] M. Morgan and D. Nathwani, “Facial palsy and infection: the unfolding story,” *Clin. Infect. Dis. Off. Publ. Infect. Dis. Soc. Am.*, vol. 14, no. 1, pp. 263–271, Jan. 1992, doi: 10.1093/clinids/14.1.263.
[24] T.-A. N. Melvin and C. J. Limb, “Overview of facial paralysis: current concepts,” *Facial Plast. Surg. FPS*, vol. 24, no. 2, pp. 155–163, May 2008, doi: 10.1055/s-2008-1075830.
[25] M. Z. Siemionow *et al.*, “First U.S. near-total human face transplantation: A paradigm shift for massive complex injuries,” *Plast. Reconstr. Surg.*, vol. 125, no. 1, pp. 111–122, 2010, doi: 10.1097/PRS.0b013e3181c15c4c.
[26] Laurent Lantieri, “Jérôme Hamon: Frenchman gets ‘third face’ in new transplant.” [Online]. Available: https://www.bbc.com/news/world-europe-43794916
[27] S. Khalifian *et al.*, “Facial transplantation: The first 9 years,” *The Lancet*, vol. 384, no. 9960, pp. 2153–2163, 2014, doi: 10.1016/S0140-6736(13)62632-X.
[28] M. Sosin and E. D. Rodriguez, “The Face Transplantation Update: 2016,” *Plast. Reconstr. Surg.*, vol. 137, no. 6, pp. 1841–1850, Jun. 2016, doi: 10.1097/PRS.0000000000002149.
[29] P. L. Dixon, X. Zhang, M. Domalain, A. M. Flores, and V. W.-H. Lin, “Physical Medicine and Rehabilitation after Face Transplantation,” in *The Know-How of Face Transplantation*, M. Z. Siemionow, Ed. London: Springer London, 2011, pp. 151–172. doi: 10.1007/978-0-85729-253-7_14.
[30] W. J. Rifkin *et al.*, “Achievements and Challenges in Facial Transplantation,” *Ann. Surg.*, vol. 268, no. 2, pp. 260–270, Aug. 2018, doi: 10.1097/SLA.0000000000002723.
[31] K. Shanmugarajah, S. Hettiaratchy, A. Clarke, and P. E. M. Butler, “Clinical outcomes of facial transplantation: A review,” *Int. J. Surg.*, vol. 9, no. 8, pp. 600–607, 2011, doi: 10.1016/j.ijsu.2011.09.005.
[32] B. Pomahac *et al.*, “Three patients with full facial transplantation,” *N. Engl. J. Med.*, vol. 366, no. 8, pp. 715–722, Feb. 2012, doi: 10.1056/NEJMoa1111432.
[33] Robinson, Mara Wernick, “Facial rehabilitation: evaluation and treatment strategies for the patient with facial palsy”, doi: 10.1016/j.otc.2018.07.011.

[34] J. G. Neely, N. G. Cherian, C. B. Dickerson, and J. M. Nedzelski, “Sunnybrook facial grading system: reliability and criteria for grading,” *The Laryngoscope*, vol. 120, no. 5, pp. 1038–1045, May 2010, doi: 10.1002/lary.20868.

[35] J. W. House and D. E. Brackmann, “Facial nerve grading system,” *Otolaryngol.--Head Neck Surg. Off. J. Am. Acad. Otolaryngol.-Head Neck Surg.*, vol. 93, no. 2, pp. 146–147, Apr. 1985, doi: 10.1177/019459988509300202.

[36] B. G. Ross, G. Fradet, and J. M. Nedzelski, “Development of a sensitive clinical facial grading system,” *Otolaryngol.--Head Neck Surg. Off. J. Am. Acad. Otolaryngol.-Head Neck Surg.*, vol. 114, no. 3, pp. 380–386, Mar. 1996, doi: 10.1016/s0194-5998(96)70206-1.

[37] J. B. Kahn, R. E. Gliklich, K. P. Boyev, M. G. Stewart, R. B. Metson, and M. J. McKenna, “Validation of a patient-graded instrument for facial nerve paralysis: the FaCE scale,” *The Laryngoscope*, vol. 111, no. 3, pp. 387–398, Mar. 2001, doi: 10.1097/00005537-200103000-00005.

[38] J. M. VanSwearingen and J. S. Brach, “The Facial Disability Index: reliability and validity of a disability assessment instrument for disorders of the facial neuromuscular system,” *Phys. Ther.*, vol. 76, no. 12, pp. 1288–1298; discussion 1298-1300, Dec. 1996, doi: 10.1093/ptj/76.12.1288.

[39] W. S. W. Samsudin and K. Sundaraj, “Clinical and non-clinical initial assessment of facial nerve paralysis: A qualitative review,” *Biocybern. Biomed. Eng.*, vol. 34, no. 2, pp. 71–78, 2014, doi: 10.1016/j.bbe.2014.02.005.

[40] D. Haase, L. Minnigerode, G. F. Volk, J. Denzler, and O. Guntinas-Lichius, “Automated and objective action coding of facial expressions in patients with acute facial palsy,” *Eur. Arch. Oto-Rhino-Laryngol. Off. J. Eur. Fed. Oto-Rhino-Laryngol. Soc. EUFOS Affil. Ger. Soc. Oto-Rhino-Laryngol. - Head Neck Surg.*, vol. 272, no. 5, pp. 1259–1267, May 2015, doi: 10.1007/s00405-014-3385-8.

[41] John McCarthy, Marvin L. Minsky, Nathaniel Rochester, and Claude E. Shannon, “A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence,” presented at the 27(4), 12, Aug. 1955. doi: https://doi.org/10.1609/aimag.v27i4.1904.

[42] E. Alpaydin, *Introduction to machine learning*, 3rd ed. Cambridge: MIT press, 2014.

[43] R. Y. Choi, A. S. Coyner, J. Kalpathy-Cramer, M. F. Chiang, and J. P. Campbell, “Introduction to Machine Learning, Neural Networks, and Deep Learning,” *Transl. Vis. Sci. Technol.*, vol. 9, no. 2, p. 14, Feb. 2020, doi: 10.1167/tvst.9.2.14.

[44] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. Cambridge, Massachusetts London, England: The MIT Press, 2016.

[45] S.-C. Wang, “Artificial Neural Network,” in *Interdisciplinary Computing in Java Programming*, Boston, MA: Springer US, 2003, pp. 81–100. doi: 10.1007/978-1-4615-0377-4_5.

[46] D. Forsyth and J. Ponce, *Computer vision: a modern approach*, 2. ed. Upper Saddle River, N.J: Prentice Hall, 2012.

[47] C. Saharia, J. Ho, W. Chan, T. Salimans, D. J. Fleet, and M. Norouzi, “Image Super-Resolution via Iterative Refinement,” *ArXiv210407636 Cs Eess*, Jun. 2021, Accessed: Apr. 14, 2022. [Online]. Available: http://arxiv.org/abs/2104.07636

[48] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, “DeepFace: Closing the Gap to Human-Level Performance in Face Verification,” in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, OH, USA, Jun. 2014, pp. 1701–1708. doi: 10.1109/CVPR.2014.220.

[49] F. Schroff, D. Kalenichenko, and J. Philbin, “FaceNet: A Unified Embedding for Face Recognition and Clustering,” *2015 IEEE Conf. Comput. Vis. Pattern Recognit. CVPR*, pp. 815–823, Jun. 2015, doi: 10.1109/CVPR.2015.7298682.

[50] H. Wang, N. Wang, and D.-Y. Yeung, “Collaborative Deep Learning for Recommender Systems,” *ArXiv14092944 Cs Stat*, Jun. 2015, Accessed: Feb. 11, 2022. [Online]. Available: http://arxiv.org/abs/1409.2944
[51] C. A. Gomez-Uribe and N. Hunt, “The Netflix Recommender System: Algorithms, Business Value, and Innovation,” *ACM Trans. Manag. Inf. Syst.*, vol. 6, no. 4, pp. 1–19, Jan. 2016, doi: 10.1145/2843948.
[52] M. Johnson *et al.*, “Google’s Multilingual Neural Machine Translation System: Enabling Zero-Shot Translation,” *ArXiv161104558 Cs*, Aug. 2017, Accessed: Feb. 11, 2022. [Online]. Available: http://arxiv.org/abs/1611.04558
[53] Lewis, William, “Skype Translator: Breaking down language and hearing barriers. A behind the scenes look at near real-time speech translation,” 2015, p. 37.
[54] A. A. Ariyo, A. O. Adewumi, and C. K. Ayo, “Stock Price Prediction Using the ARIMA Model,” in *2014 UKSim-AMSS 16th International Conference on Computer Modelling and Simulation*, Cambridge, United Kingdom, Mar. 2014, pp. 106–112. doi: 10.1109/UKSim.2014.67.
[55] C. K.-S. Leung, R. K. MacKinnon, and Y. Wang, “A machine learning approach for stock price prediction,” in *Proceedings of the 18th International Database Engineering & Applications Symposium on - IDEAS ’14*, Porto, Portugal, 2014, pp. 274–277. doi: 10.1145/2628194.2628211.
[56] A. Roy, J. Sun, R. Mahoney, L. Alonzi, S. Adams, and P. Beling, “Deep learning detecting fraud in credit card transactions,” in *2018 Systems and Information Engineering Design Symposium (SIEDS)*, Charlottesville, VA, Apr. 2018, pp. 129–134. doi: 10.1109/SIEDS.2018.8374722.
[57] M. U. Gudelek, S. A. Boluk, and A. M. Ozbayoglu, “A deep learning based stock trading model with 2-D CNN trend detection,” in *2017 IEEE Symposium Series on Computational Intelligence (SSCI)*, Honolulu, HI, Nov. 2017, pp. 1–8. doi: 10.1109/SSCI.2017.8285188.
[58] M. Sameen and B. Pradhan, “Severity Prediction of Traffic Accidents with Recurrent Neural Networks,” *Appl. Sci.*, vol. 7, no. 6, p. 476, Jun. 2017, doi: 10.3390/app7060476.
[59] X. Ma, H. Yu, Y. Wang, and Y. Wang, “Large-Scale Transportation Network Congestion Evolution Prediction Using Deep Learning Theory,” *PLOS ONE*, vol. 10, no. 3, p. e0119044, Mar. 2015, doi: 10.1371/journal.pone.0119044.
[60] EVAN ACKERMAN and ALYSSA PAGANO, “Deep-Learning First: Drive.ai’s Path to Autonomous Driving,” spectrum, Mar. 13, 2017.
[61] D. Cireşan, U. Meier, J. Masci, and J. Schmidhuber, “Multi-column deep neural network for traffic sign classification,” *Neural Netw. Off. J. Int. Neural Netw. Soc.*, vol. 32, pp. 333–338, Aug. 2012, doi: 10.1016/j.neunet.2012.02.023.
[62] Z. Obermeyer and E. J. Emanuel, “Predicting the Future - Big Data, Machine Learning, and Clinical Medicine,” *N. Engl. J. Med.*, vol. 375, no. 13, pp. 1216–1219, Sep. 2016, doi: 10.1056/NEJMp1606181.
[63] D. S. Watson *et al.*, “Clinical applications of machine learning algorithms: beyond the black box,” *BMJ*, vol. 364, p. l886, Mar. 2019, doi: 10.1136/bmj.l886.
[64] V. Gulshan *et al.*, “Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs,” *JAMA*, vol. 316, no. 22, pp. 2402–2410, Dec. 2016, doi: 10.1001/jama.2016.17216.
[65] P. Rajpurkar *et al.*, “CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning,” *ArXiv171105225 Cs Stat*, Dec. 2017, Accessed: Feb. 11, 2022. [Online]. Available: http://arxiv.org/abs/1711.05225

[66] A. Esteva *et al.*, “Dermatologist-level classification of skin cancer with deep neural networks,” *Nature*, vol. 542, no. 7639, pp. 115–118, Feb. 2017, doi: 10.1038/nature21056.
[67] W. A. Kibbe *et al.*, “Disease Ontology 2015 update: an expanded and updated database of human diseases for linking biomedical knowledge through disease data,” *Nucleic Acids Res.*, vol. 43, no. Database issue, pp. D1071-1078, Jan. 2015, doi: 10.1093/nar/gku1011.
[68] S. Bannur *et al.*, “Hierarchical Analysis of Visual COVID-19 Features from Chest Radiographs,” *ArXiv210706618 Cs Eess*, Jul. 2021, Accessed: Apr. 16, 2022. [Online]. Available: http://arxiv.org/abs/2107.06618
[69] J. De Fauw *et al.*, “Clinically applicable deep learning for diagnosis and referral in retinal disease,” *Nat. Med.*, vol. 24, no. 9, pp. 1342–1350, Sep. 2018, doi: 10.1038/s41591-018-0107-6.
[70] S. P. Somashekhar *et al.*, “Watson for Oncology and breast cancer treatment recommendations: agreement with an expert multidisciplinary tumor board,” *Ann. Oncol. Off. J. Eur. Soc. Med. Oncol.*, vol. 29, no. 2, pp. 418–423, Feb. 2018, doi: 10.1093/annonc/mdx781.
[71] Gandhia, Neel and Shakti Mishraa, “Applications of Reinforcement learning for Medical Decision Making.” 2021.
[72] Z. Zhang and written on behalf of AME Big-Data Clinical Trial Collaborative Group, “Reinforcement learning in clinical medicine: a method to optimize dynamic treatment regime over time,” *Ann. Transl. Med.*, vol. 7, no. 14, p. 345, Jul. 2019, doi: 10.21037/atm.2019.06.75.
[73] D. Ernst, G.-B. Stan, J. Goncalves, and L. Wehenkel, “Clinical data based optimal STI strategies for HIV: a reinforcement learning approach,” in *Proceedings of the 45th IEEE Conference on Decision and Control*, San Diego, CA, Dec. 2006, pp. 667–672. doi: 10.1109/CDC.2006.377527.
[74] E. Yom-Tov, G. Feraru, M. Kozdoba, S. Mannor, M. Tennenholtz, and I. Hochberg, “Encouraging Physical Activity in Patients With Diabetes: Intervention Using a Reinforcement Learning System,” *J. Med. Internet Res.*, vol. 19, no. 10, p. e338, Oct. 2017, doi: 10.2196/jmir.7994.
[75] Y. Zhao, M. R. Kosorok, and D. Zeng, “Reinforcement learning design for cancer clinical trials,” *Stat. Med.*, vol. 28, no. 26, pp. 3294–3315, Nov. 2009, doi: 10.1002/sim.3720.
[76] F. Sahba, H. R. Tizhoosh, and M. M. A. Salama, “Application of reinforcement learning for segmentation of transrectal ultrasound images,” *BMC Med. Imaging*, vol. 8, p. 8, Apr. 2008, doi: 10.1186/1471-2342-8-8.
[77] Yuan Ling *et al.*, “Diagnostic inferencing via improving clinical concept extraction with deep reinforcement learning: A preliminary study,” presented at the Machine Learning for Healthcare Conference, 2017.
[78] E. M. Shakshuki, M. Reid, and T. R. Sheltami, “An Adaptive User Interface in Healthcare,” *Procedia Comput. Sci.*, vol. 56, pp. 49–58, 2015, doi: 10.1016/j.procs.2015.07.182.
[79] J. Mulani, S. Heda, K. Tumdi, J. Patel, H. Chhinkaniwala, and J. Patel, “Deep Reinforcement Learning Based Personalized Health Recommendations,” in *Deep Learning Techniques for Biomedical and Health Informatics*, vol. 68, S. Dash, B. R. Acharya, M. Mittal, A. Abraham, and A. Kelemen, Eds. Cham: Springer International Publishing, 2020, pp. 231–255. doi: 10.1007/978-3-030-33966-1_12.
[80] R. T. Sutton, D. Pincock, D. C. Baumgart, D. C. Sadowski, R. N. Fedorak, and K. I. Kroeker, “An overview of clinical decision support systems: benefits, risks, and

strategies for success," *Npj Digit. Med.*, vol. 3, no. 1, p. 17, Dec. 2020, doi: 10.1038/s41746-020-0221-y.
[81] V. Wattanasoontorn, I. Boada, R. García, and M. Sbert, "Serious games for health," *Entertain. Comput.*, vol. 4, no. 4, pp. 231–247, Dec. 2013, doi: 10.1016/j.entcom.2013.09.002.
[82] J.-A. Kim, K.-K. Kang, H.-R. Yang, and D. Kim, "A Sensory Gate-Ball Game for the Aged People and its User Interface Design," in *2009 Conference in Games and Virtual Worlds for Serious Applications*, Coventry, UK, Mar. 2009, pp. 111–116. doi: 10.1109/VS-GAMES.2009.19.
[83] A. Laikari, "Exergaming - Gaming for health: A bridge between real world and virtual communities," in *2009 IEEE 13th International Symposium on Consumer Electronics*, Kyoto, Japan, May 2009, pp. 665–668. doi: 10.1109/ISCE.2009.5157004.
[84] XIAO Cheng-yong and Xiang Wei-ming, "Constructing 3d game engine based on xna," presented at the Computer Knowledge and Technology, 2010.
[85] A. Väätänen and J. Leikas, "Human-Centred Design and Exercise Games," in *Design and Use of Serious Games*, vol. 37, M. Kankaanranta and P. Neittaanmäki, Eds. Dordrecht: Springer Netherlands, 2009, pp. 33–47. doi: 10.1007/978-1-4020-9496-5_3.
[86] "Medical College of Georgia School of Dentistry faculty and students." Accessed: Feb. 22, 2022. [Online]. Available: https://medicalxpress.com/news/2009-06-simulation-students-dental-implant-procedures.html
[87] "Applied Research Associates, Inc. (2012) HumanSim: a high-fidelity virtual hospital." Accessed: Feb. 22, 2022. [Online]. Available: https://apps.apple.com/us/app/humansim-preview/id437066468
[88] M. Vazquez-Vazquez, V. Santana-Lopez, M. Skodova, J. Ferrero-Alvarez-Rementeria, and A. Torres-Olivera, "Hand hygiene training through a serious game: New ways of improving Safe Practices," in *2011 IEEE 1st International Conference on Serious Games and Applications for Health (SeGAH)*, Braga, Portugal, Nov. 2011, pp. 1–2. doi: 10.1109/SeGAH.2011.6165439.
[89] Jesús Reseco Gago, Teresa Meneu Barreira, Rebeca González Carrascosa, and Purificación García Segovia, "Nutritional Serious-Games platform," presented at the eChallenges e-2010 Conference, 2010.
[90] S. Lee, J. Kim, J. Kim, and M. Lee, "A design of the u-health monitoring system using a Nintendo DS game machine," *Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. IEEE Eng. Med. Biol. Soc. Annu. Int. Conf.*, vol. 2009, pp. 1695–1698, 2009, doi: 10.1109/IEMBS.2009.5333902.
[91] P. Fergus, K. Kifayat, S. Cooper, M. Merabti, and A. El Rhalibi, "A framework for physical health improvement using Wireless Sensor Networks and gaming," presented at the 3d International ICST Conference on Pervasive Computing Technologies for Healthcare, London, UK, 2009. doi: 10.4108/ICST.PERVASIVEHEALTH2009.6045.
[92] A. De Bortoli and O. Gaggi, "PlayWithEyes: A new way to test children eyes," in *2011 IEEE 1st International Conference on Serious Games and Applications for Health (SeGAH)*, Braga, Portugal, Nov. 2011, pp. 1–4. doi: 10.1109/SeGAH.2011.6165458.
[93] J. A. McKanna, H. Jimison, and M. Pavel, "Divided attention in computer game play: analysis utilizing unobtrusive health monitoring," *Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. IEEE Eng. Med. Biol. Soc. Annu. Int. Conf.*, vol. 2009, pp. 6247–6250, 2009, doi: 10.1109/IEMBS.2009.5334662.

[94] Q. Wang, O. Sourina, and M. K. Nguyen, “EEG-Based ‘Serious’ Games Design for Medical Applications,” in *2010 International Conference on Cyberworlds*, Singapore, Singapore, Oct. 2010, pp. 270–276. doi: 10.1109/CW.2010.56.
[95] S. D. Atkinson and V. L. Narasimhan, “Design of an introductory medical gaming environment for diagnosis and management of Parkinson’s disease,” in *Trendz in Information Sciences & Computing(TISC2010)*, Chennai, India, Dec. 2010, pp. 94–102. doi: 10.1109/TISC.2010.5714615.
[96] M. Cagatay, P. Ege, G. Tokdemir, and N. E. Cagiltay, “A serious game for speech disorder children therapy,” in *2012 7th International Symposium on Health Informatics and Bioinformatics*, Nevsehir, Turkey, Apr. 2012, pp. 18–23. doi: 10.1109/HIBIT.2012.6209036.
[97] M. Pedraza-Hueso, S. Martín-Calzón, F. J. Díaz-Pernas, and M. Martínez-Zarzuela, “Rehabilitation Using Kinect-based Games and Virtual Reality,” *Procedia Comput. Sci.*, vol. 75, no. Vare, pp. 161–168, 2015, doi: 10.1016/j.procs.2015.12.233.
[98] H. Tannous, D. Istrate, M. C. Ho Ba Tho, and T. T. Dao, “Feasibility study of a serious game based on Kinect system for functional rehabilitation of the lower limbs,” *Eur. Res. Telemed. Rech. Eur. En Télémédecine*, vol. 5, no. 3, pp. 97–104, Sep. 2016, doi: 10.1016/j.eurtel.2016.05.004.
[99] S. S. Esfahlani, B. Muresan, A. Sanaei, and G. Wilson, “Validity of the Kinect and Myo armband in a serious game for assessing upper limb movement,” *Entertain. Comput.*, vol. 27, no. August 2017, pp. 150–156, 2018, doi: 10.1016/j.entcom.2018.05.003.
[100] V. Fernandez-Cervantes, N. Neubauer, B. Hunter, E. Stroulia, and L. Liu, “VirtualGym: A kinect-based system for seniors exercising at home,” *Entertain. Comput.*, vol. 27, no. March, pp. 60–72, 2018, doi: 10.1016/j.entcom.2018.04.001.
[101] M. Xu *et al.*, “Personalized training through Kinect-based games for physical education,” *J. Vis. Commun. Image Represent.*, vol. 62, pp. 394–401, 2019, doi: 10.1016/j.jvcir.2019.05.007.
[102] R. Cabrera, A. Molina, I. Gómez, and J. García-Heras, “Kinect as an access device for people with cerebral palsy: A preliminary study,” *Int. J. Hum. Comput. Stud.*, vol. 108, no. July, pp. 62–69, 2017, doi: 10.1016/j.ijhcs.2017.07.004.
[103] H. Nauta and T. A. M. Spil, “Change your lifestyle or your game is over: The design of a serious game for diabetes,” in *2011 IEEE 1st International Conference on Serious Games and Applications for Health (SeGAH)*, Braga, Portugal, Nov. 2011, pp. 1–7. doi: 10.1109/SeGAH.2011.6165436.
[104] D. González-Ortega, F. J. Díaz-Pernas, M. Martínez-Zarzuela, and M. Antón-Rodríguez, “A Kinect-based system for cognitive rehabilitation exercises monitoring,” *Comput. Methods Programs Biomed.*, vol. 113, no. 2, pp. 620–631, Feb. 2014, doi: 10.1016/j.cmpb.2013.10.014.
[105] T.-N. Nguyen, “Clinical decision support system for facial mimic rehabilitation,”, Ph.D. dissertation, Université de Technologie de Compiègne, 2020.
[106] M.-C. Ho Ba Tho and T. T. Dao, “Knowledge Extraction from Medical Imaging for Advanced Patient-Specific Musculoskeletal Models,” in *Encyclopedia of Biomedical Engineering*, Elsevier, 2019, pp. 135–142. doi: 10.1016/B978-0-12-801238-3.99935-5.
[107] T. Cawthorne and D. R. Haynes, “Facial Palsy,” *BMJ*, vol. 2, no. 5003, pp. 1197–1200, Nov. 1956, doi: 10.1136/bmj.2.5003.1197.
[108] Hotton, “The psychosocial impact of facial palsy: A systematic review,” *Br. J. Health Psychol.*, doi: 10.1111/bjhp.12440.
[109] T.-N. Nguyen, S. Dakpe, M.-C. Ho Ba Tho, and T.-T. Dao, “Kinect-driven Patient-specific Head, Skull, and Muscle Network Modelling for Facial Palsy Patients,”

Comput. Methods Programs Biomed., vol. 200, p. 105846, Mar. 2021, doi: 10.1016/j.cmpb.2020.105846.
[110] T.-N. Nguyen, V.-D. Tran, H.-Q. Nguyen, D.-P. Nguyen, and T.-T. Dao, "Enhanced head-skull shape learning using statistical modeling and topological features," *Med. Biol. Eng. Comput.*, Jan. 2022, doi: 10.1007/s11517-021-02483-y.
[111] T.-N. Nguyen, S. Dakpé, M.-C. Ho Ba Tho, and T.-T. Dao, "Real-time computer vision system for tracking simultaneously subject-specific rigid head and non-rigid facial mimic movements using a contactless sensor and system of systems approach," *Comput. Methods Programs Biomed.*, vol. 191, p. 105410, Jul. 2020, doi: 10.1016/j.cmpb.2020.105410.
[112] Lijun Yin, Xiaozhou Wei, Yi Sun, Jun Wang, and M. J. Rosato, "A 3D Facial Expression Database for Facial Behavior Research," in *7th International Conference on Automatic Face and Gesture Recognition (FGR06)*, Southampton, UK, 2006, pp. 211–216. doi: 10.1109/FGR.2006.6.
[113] A. Savran *et al.*, "Bosphorus Database for 3D Face Analysis," in *Biometrics and Identity Management*, Berlin, Heidelberg, 2008, pp. 47–56.
[114] Zhang, "Intelligent facial emotion recognition and semantic-based topic detection for a humanoid robot", doi: https://doi.org/10.1016/j.eswa.2013.03.016.
[115] F. Dornaika and B. Raducanu, "Efficient Facial Expression Recognition for Human Robot Interaction," in *Computational and Ambient Intelligence*, vol. 4507, F. Sandoval, A. Prieto, J. Cabestany, and M. Graña, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, pp. 700–708. doi: 10.1007/978-3-540-73007-1_84.
[116] T. Weise, S. Bouaziz, H. Li, and M. Pauly, "Realtime performance-based facial animation," in *ACM SIGGRAPH 2011 papers on - SIGGRAPH '11*, Vancouver, British Columbia, Canada, 2011, p. 1. doi: 10.1145/1964921.1964972.
[117] Y. Lee, D. Terzopoulos, and K. Walters, "Realistic modeling for facial animation," in *Proceedings of the 22nd annual conference on Computer graphics and interactive techniques - SIGGRAPH '95*, Not Known, 1995, pp. 55–62. doi: 10.1145/218380.218407.
[118] T. Weise, H. Li, L. Van Gool, and M. Pauly, "Face/Off: live facial puppetry," in *Proceedings of the 2009 ACM SIGGRAPH/Eurographics Symposium on Computer Animation - SCA '09*, New Orleans, Louisiana, 2009, p. 7. doi: 10.1145/1599470.1599472.
[119] M. Leo, P. Carcagnì, P. L. Mazzeo, P. Spagnolo, D. Cazzato, and C. Distante, "Analysis of Facial Information for Healthcare Applications: A Survey on Computer Vision-Based Approaches," *Information*, vol. 11, no. 3, p. 128, Feb. 2020, doi: 10.3390/info11030128.
[120] M. C. EL Rai, N. Werghi, H. Al Muhairi, and H. Alsafar, "Using facial images for the diagnosis of genetic syndromes: A survey," in *2015 International Conference on Communications, Signal Processing, and their Applications (ICCSPA'15)*, Sharjah, United Arab Emirates, Feb. 2015, pp. 1–6. doi: 10.1109/ICCSPA.2015.7081271.
[121] A. Kermi, S. Marniche-Kermi, and M. T. Laskri, "3D-Computerized facial reconstructions from 3D-MRI of human heads using deformable model approach," in *2010 International Conference on Machine and Web Intelligence*, Algiers, Algeria, Oct. 2010, pp. 276–282. doi: 10.1109/ICMWI.2010.5648144.
[122] C. Flynn, I. Stavness, J. Lloyd, and S. Fels, "A finite element model of the face including an orthotropic skin model under *in vivo* tension," *Comput. Methods Biomech. Biomed. Engin.*, vol. 18, no. 6, pp. 571–582, Apr. 2015, doi: 10.1080/10255842.2013.820720.

[123] T. Beeler, B. Bickel, P. Beardsley, B. Sumner, and M. Gross, “High-quality single-shot capture of facial geometry,” in *ACM SIGGRAPH 2010 papers on - SIGGRAPH ’10*, Los Angeles, California, 2010, p. 1. doi: 10.1145/1833349.1778777.
[124] C.-H. Chen, I.-J. Lee, and L.-Y. Lin, “Augmented reality-based self-facial modeling to promote the emotional expression and social skills of adolescents with autism spectrum disorders,” *Res. Dev. Disabil.*, vol. 36, pp. 396–403, Jan. 2015, doi: 10.1016/j.ridd.2014.10.015.
[125] C. Li, A. Barreto, C. Chin, and J. Zhai, “Biometric identification using 3D face scans,” *Biomed. Sci. Instrum.*, vol. 42, pp. 320–325, 2006.
[126] D. Kim, M. Hernandez, J. Choi, and G. Medioni, “Deep 3D Face Identification,” *ArXiv170310714 Cs*, Mar. 2017, Accessed: Jan. 13, 2022. [Online]. Available: http://arxiv.org/abs/1703.10714
[127] D.-P. Nguyen, M.-C. Ho Ba Tho, and T.-T. Dao, “Enhanced facial expression recognition using 3D point sets and geometric deep learning,” *Med. Biol. Eng. Comput.*, vol. 59, no. 6, pp. 1235–1244, Jun. 2021, doi: 10.1007/s11517-021-02383-1.
[128] A. Morales, G. Piella, and F. M. Sukno, “Survey on 3D face reconstruction from uncalibrated images,” *ArXiv201105740 Cs*, Feb. 2021, Accessed: Jan. 13, 2022. [Online]. Available: http://arxiv.org/abs/2011.05740
[129] A. Bas, W. A. P. Smith, T. Bolkart, and S. Wuhrer, “Fitting a 3D Morphable Model to Edges: A Comparison Between Hard and Soft Correspondences,” in *Computer Vision – ACCV 2016 Workshops*, vol. 10117, C.-S. Chen, J. Lu, and K.-K. Ma, Eds. Cham: Springer International Publishing, 2017, pp. 377–391. doi: 10.1007/978-3-319-54427-4_28.
[130] Xiangyu Zhu, Junjie Yan, Dong Yi, Zhen Lei, and S. Z. Li, “Discriminative 3D morphable model fitting,” in *2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*, Ljubljana, May 2015, pp. 1–8. doi: 10.1109/FG.2015.7163096.
[131] O. Aldrian and W. A. P. Smith, “Inverse Rendering of Faces with a 3D Morphable Model,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 5, pp. 1080–1093, May 2013, doi: 10.1109/TPAMI.2012.206.
[132] V. Blanz and T. Vetter, “A morphable model for the synthesis of 3D faces,” in *Proceedings of the 26th annual conference on Computer graphics and interactive techniques - SIGGRAPH*, 1999, pp. 187–194. doi: 10.1145/311535.311556.
[133] P. Paysan, R. Knothe, B. Amberg, S. Romdhani, and T. Vetter, “A 3D Face Model for Pose and Illumination Invariant Face Recognition,” in *2009 Sixth IEEE International Conference on Advanced Video and Signal Based Surveillance*, Genova, Italy, Sep. 2009, pp. 296–301. doi: 10.1109/AVSS.2009.58.
[134] P. Huber *et al.*, “A Multiresolution 3D Morphable Face Model and Fitting Framework,” in *Proceedings of the 11th Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*, Rome, Italy, 2016, pp. 79–86. doi: 10.5220/0005669500790086.
[135] J. Booth, A. Roussos, A. Ponniah, D. Dunaway, and S. Zafeiriou, “Large Scale 3D Morphable Models,” *Int. J. Comput. Vis.*, vol. 126, no. 2–4, pp. 233–254, Apr. 2018, doi: 10.1007/s11263-017-1009-7.
[136] H. Dai, N. Pears, W. Smith, and C. Duncan, “A 3D Morphable Model of Craniofacial Shape and Texture Variation,” in *2017 IEEE International Conference on Computer Vision (ICCV)*, Venice, Oct. 2017, pp. 3104–3112. doi: 10.1109/ICCV.2017.335.
[137] H. Dai, N. Pears, W. Smith, and C. Duncan, “Statistical Modeling of Craniofacial Shape and Texture,” *Int. J. Comput. Vis.*, vol. 128, no. 2, pp. 547–571, Feb. 2020, doi: 10.1007/s11263-019-01260-7.

[138] Chen Cao, Yanlin Weng, Shun Zhou, Yiying Tong, and Kun Zhou, “FaceWarehouse: A 3D Facial Expression Database for Visual Computing,” *IEEE Trans. Vis. Comput. Graph.*, vol. 20, no. 3, pp. 413–425, Mar. 2014, doi: 10.1109/TVCG.2013.249.
[139] I. Kemelmacher-Shlizerman and S. M. Seitz, “Face reconstruction in the wild,” in *2011 International Conference on Computer Vision*, Barcelona, Spain, Nov. 2011, pp. 1746–1753. doi: 10.1109/ICCV.2011.6126439.
[140] S. Suwajanakorn, I. Kemelmacher-Shlizerman, and S. M. Seitz, “Total Moving Face Reconstruction,” in *Computer Vision – ECCV 2014*, vol. 8692, D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, Eds. Cham: Springer International Publishing, 2014, pp. 796–812. doi: 10.1007/978-3-319-10593-2_52.
[141] P. Snape, Y. Panagakis, and S. Zafeiriou, “Automatic construction of robust spherical harmonic subspaces,” in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, USA, Jun. 2015, pp. 91–100. doi: 10.1109/CVPR.2015.7298604.
[142] D. Zeng, Q. Zhao, S. Long, and J. Li, “Examplar coherent 3D face reconstruction from forensic mugshot database,” *Image Vis. Comput.*, vol. 58, pp. 193–203, Feb. 2017, doi: 10.1016/j.imavis.2016.03.001.
[143] I. Kemelmacher-Shlizerman and R. Basri, “3D Face Reconstruction from a Single Image Using a Single Reference Face Shape,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 2, pp. 394–405, Feb. 2011, doi: 10.1109/TPAMI.2010.63.
[144] Mingli Song, Dacheng Tao, Xiaoqin Huang, Chun Chen, and Jiajun Bu, “Three-Dimensional Face Reconstruction from a Single Image by a Coupled RBF Network,” *IEEE Trans. Image Process.*, vol. 21, no. 5, pp. 2887–2897, May 2012, doi: 10.1109/TIP.2012.2183882.
[145] M. Lee and C.-H. Choi, “Fast facial shape recovery from a single image with general, unknown lighting by using tensor representation,” *Pattern Recognit.*, vol. 44, no. 7, pp. 1487–1496, Jul. 2011, doi: 10.1016/j.patcog.2010.12.018.
[146] M. Lee and C.-H. Choi, “A robust real-time algorithm for facial shape recovery from a single image containing cast shadow under general, unknown lighting,” *Pattern Recognit.*, vol. 46, no. 1, pp. 38–44, Jan. 2013, doi: 10.1016/j.patcog.2012.06.016.
[147] M. Lee and C.-H. Choi, “Real-time facial shape recovery from a single image under general, unknown lighting by rank relaxation,” *Comput. Vis. Image Underst.*, vol. 120, pp. 59–69, Mar. 2014, doi: 10.1016/j.cviu.2013.12.010.
[148] X. Cao, Z. Chen, A. Chen, X. Chen, S. Li, and J. Yu, “Sparse Photometric 3D Face Reconstruction Guided by Morphable Models,” in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, Jun. 2018, pp. 4635–4644. doi: 10.1109/CVPR.2018.00487.
[149] Y. Li, L. Ma, H. Fan, and K. Mitchell, “Feature-preserving detailed 3D face reconstruction from a single image,” in *Proceedings of the 15th ACM SIGGRAPH European Conference on Visual Media Production - CVMP ’18*, London, United Kingdom, 2018, pp. 1–9. doi: 10.1145/3278471.3278473.
[150] G. Rotger, F. Moreno-Noguer, F. Lumbreras, and A. Agudo, “Detailed 3D Face Reconstruction from a Single RGB Image,” *J. WSCG*, vol. 27, no. 2, 2019, doi: 10.24132/JWSCG.2019.27.2.3.
[151] G. Zhang, H. Han, S. Shan, X. Song, and X. Chen, “Face Alignment across Large Pose via MT-CNN Based 3D Shape Reconstruction,” in *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*, Xi’an, May 2018, pp. 210–217. doi: 10.1109/FG.2018.00039.
[152] Y. Zhou, J. Deng, I. Kotsia, and S. Zafeiriou, “Dense 3D Face Decoding over 2500FPS: Joint Texture & Shape Convolutional Mesh Decoders,” *ArXiv190403525*

Cs, Apr. 2019, Accessed: Jan. 13, 2022. [Online]. Available: http://arxiv.org/abs/1904.03525
[153] X. Zhu, Z. Lei, X. Liu, H. Shi, and S. Z. Li, "Face Alignment Across Large Poses: A 3D Solution," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 1, pp. 78–92, Jan. 2019, doi: 10.1109/TPAMI.2017.2778152.
[154] L. Galteri, C. Ferrari, G. Lisanti, S. Berretti, and A. D. Bimbo, "Coarse to Fine 3D Face Reconstruction from Single Image," in *2019 14th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2019)*, Lille, France, May 2019, pp. 1–1. doi: 10.1109/FG.2019.8756603.
[155] L. Galteri, C. Ferrari, G. Lisanti, S. Berretti, and A. Del Bimbo, "Deep 3D morphable model refinement via progressive growing of conditional Generative Adversarial Networks," *Comput. Vis. Image Underst.*, vol. 185, pp. 31–42, Aug. 2019, doi: 10.1016/j.cviu.2019.05.002.
[156] E. Richardson, M. Sela, and R. Kimmel, "3D Face Reconstruction by Learning from Synthetic Data," *ArXiv160904387 Cs*, Sep. 2016, Accessed: Feb. 11, 2022. [Online]. Available: http://arxiv.org/abs/1609.04387
[157] J. Piao, C. Qian, and H. Li, "Semi-Supervised Monocular 3D Face Reconstruction with End-to-End Shape-Preserved Domain Transfer," in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, Seoul, Korea (South), Oct. 2019, pp. 9397–9406. doi: 10.1109/ICCV.2019.00949.
[158] A. Tewari *et al.*, "MoFA: Model-based Deep Convolutional Face Autoencoder for Unsupervised Monocular Reconstruction," *ArXiv170310580 Cs*, Dec. 2017, Accessed: Feb. 11, 2022. [Online]. Available: http://arxiv.org/abs/1703.10580
[159] L. Tran and X. Liu, "Nonlinear 3D Face Morphable Model," *ArXiv180403786 Cs*, Aug. 2018, Accessed: Feb. 11, 2022. [Online]. Available: http://arxiv.org/abs/1804.03786
[160] L. Tran and X. Liu, "On Learning 3D Face Morphable Model from In-the-wild Images," *IEEE Trans. Pattern Anal. Mach. Intell.*, pp. 1–1, 2019, doi: 10.1109/TPAMI.2019.2927975.
[161] X. Chai, J. Chen, C. Liang, D. Xu, and C.-W. Lin, "Expression-Aware Face Reconstruction Via A Dual-Stream Network," in *2020 IEEE International Conference on Multimedia and Expo (ICME)*, London, United Kingdom, Jul. 2020, pp. 1–6. doi: 10.1109/ICME46284.2020.9102811.
[162] A. Jourabloo, M. Ye, X. Liu, and L. Ren, "Pose-Invariant Face Alignment with a Single CNN," in *2017 IEEE International Conference on Computer Vision (ICCV)*, Venice, Oct. 2017, pp. 3219–3228. doi: 10.1109/ICCV.2017.347.
[163] A. Tewari *et al.*, "Self-supervised Multi-level Face Model Learning for Monocular Reconstruction at over 250 Hz," *ArXiv171202859 Cs*, Mar. 2018, Accessed: Feb. 11, 2022. [Online]. Available: http://arxiv.org/abs/1712.02859
[164] N. Savov, M. L. Ngo, S. Karaoglu, H. Dibeklioglu, and T. Gevers, "Pose and Expression Robust Age Estimation via 3D Face Reconstruction from a Single Image," in *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*, Seoul, Korea (South), Oct. 2019, pp. 1270–1278. doi: 10.1109/ICCVW.2019.00160.
[165] H. Kim, M. Zollhöfer, A. Tewari, J. Thies, C. Richardt, and C. Theobalt, "InverseFaceNet: Deep Monocular Inverse Face Rendering," *ArXiv170310956 Cs*, May 2018, Accessed: Feb. 11, 2022. [Online]. Available: http://arxiv.org/abs/1703.10956
[166] H. Yi *et al.*, "MMFace: A Multi-Metric Regression Network for Unconstrained Face Reconstruction," in *2019 IEEE/CVF Conference on Computer Vision and Pattern*

Recognition (CVPR), Long Beach, CA, USA, Jun. 2019, pp. 7655–7664. doi: 10.1109/CVPR.2019.00785.
[167] Xiangyu Zhu *et al.*, "Beyond 3DMM Space: Towards Fine-Grained 3D Face Reconstruction," presented at the ECCV.
[168] B. Chaudhuri, N. Vesdapunt, L. Shapiro, and B. Wang, "Personalized Face Modeling for Improved Face Reconstruction and Motion Retargeting," *ArXiv200706759 Cs*, Jul. 2020, Accessed: Feb. 11, 2022. [Online]. Available: http://arxiv.org/abs/2007.06759
[169] X. Li, Z. Weng, J. Liang, L. Cei, Y. Xiang, and Y. Fu, "A Novel Two-Pathway Encoder-Decoder Network for 3D Face Reconstruction," in *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Barcelona, Spain, May 2020, pp. 3682–3686. doi: 10.1109/ICASSP40776.2020.9053699.
[170] X. Tu *et al.*, "3D Face Reconstruction from A Single Image Assisted by 2D Face Images in the Wild," *IEEE Trans. Multimed.*, vol. 23, pp. 1160–1172, 2021, doi: 10.1109/TMM.2020.2993962.
[171] Z. Gao, J. Zhang, Y. Guo, C. Ma, G. Zhai, and X. Yang, "Semi-supervised 3D Face Representation Learning from Unconstrained Photo Collections," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Seattle, WA, USA, Jun. 2020, pp. 1426–1435. doi: 10.1109/CVPRW50498.2020.00182.
[172] Ayush Tewari *et al.*, "Fml: Face model learning from videos," 2019, pp. 10812--10822.
[173] C. Bhagavatula, C. Zhu, K. Luu, and M. Savvides, "Faster Than Real-time Facial Alignment: A 3D Spatial Transformer Network Approach in Unconstrained Poses," *ArXiv170705653 Cs*, Sep. 2017, Accessed: Feb. 11, 2022. [Online]. Available: http://arxiv.org/abs/1707.05653
[174] X. Fan, S. Cheng, K. Huyan, M. Hou, R. Liu, and Z. Luo, "Dual Neural Networks Coupling Data Regression with Explicit Priors for Monocular 3D Face Reconstruction," *IEEE Trans. Multimed.*, vol. 23, pp. 1252–1263, 2021, doi: 10.1109/TMM.2020.2994506.
[175] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," *ArXiv151203385 Cs*, Dec. 2015, Accessed: Jan. 13, 2022. [Online]. Available: http://arxiv.org/abs/1512.03385
[176] X. Wang, Y. Guo, B. Deng, and J. Zhang, "Lightweight Photometric Stereo for Facial Details Recovery," *ArXiv200312307 Cs*, Mar. 2020, Accessed: Feb. 11, 2022. [Online]. Available: http://arxiv.org/abs/2003.12307
[177] P. Dou, S. K. Shah, and I. A. Kakadiaris, "End-to-end 3D face reconstruction with deep neural networks," *ArXiv170405020 Cs*, Apr. 2017, Accessed: Feb. 11, 2022. [Online]. Available: http://arxiv.org/abs/1704.05020
[178] O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Deep Face Recognition," in *Procedings of the British Machine Vision Conference 2015*, Swansea, 2015, p. 41.1-41.12. doi: 10.5244/C.29.41.
[179] O. S. Ekundayo and S. Viriri, "Facial Expression Recognition: A Review of Trends and Techniques," *IEEE Access*, vol. 9, pp. 136944–136973, 2021, doi: 10.1109/ACCESS.2021.3113464.
[180] S. Hickson, N. Dufour, A. Sud, V. Kwatra, and I. Essa, "Eyemotion: Classifying Facial Expressions in VR Using Eye-Tracking Cameras," in *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*, Waikoloa Village, HI, USA, Jan. 2019, pp. 1626–1635. doi: 10.1109/WACV.2019.00178.

[181] P. Dulguerov, F. Marchal, D. Wang, and C. Gysin, “Review of objective topographic facial nerve evaluation methods,” *Am. J. Otol.*, vol. 20, no. 5, pp. 672–678, Sep. 1999.
[182] B. Sonawane and P. Sharma, “Review of automated emotion-based quantification of facial expression in Parkinson’s patients,” *Vis. Comput.*, vol. 37, no. 5, pp. 1151–1167, May 2021, doi: 10.1007/s00371-020-01859-9.
[183] A. Kołakowska, A. Landowska, M. Szwoch, W. Szwoch, and M. R. Wróbel, “Emotion Recognition and Its Applications,” in *Human-Computer Systems Interaction: Backgrounds and Applications 3*, vol. 300, Z. S. Hippe, J. L. Kulikowski, T. Mroczek, and J. Wtorek, Eds. Cham: Springer International Publishing, 2014, pp. 51–62. doi: 10.1007/978-3-319-08491-6_5.
[184] Zhou Sheng, Lin Zhu-ying, and Dong Wan-xin, “The model of E-learning based on affective computing,” in *2010 3rd International Conference on Advanced Computer Theory and Engineering (ICACTE)*, Chengdu, China, Aug. 2010, pp. V3-269-V3-272. doi: 10.1109/ICACTE.2010.5579627.
[185] C. L. Lisetti and D. J. Schiano, “Automatic facial expression interpretation: Where human-computer interaction, artificial intelligence and cognitive science intersect,” *Pragmat. Cogn.*, vol. 8, no. 1, pp. 185–235, May 2000, doi: 10.1075/pc.8.1.09lis.
[186] C. Yang *et al.*, “Different levels of facial expression recognition in patients with first-episode schizophrenia: A functional MRI study,” *Gen. Psychiatry*, vol. 31, no. 2, p. e000014, 2018, doi: 10.1136/gpsych-2018-000014.
[187] S. Poria, A. Mondal, and P. Mukhopadhyay, “Evaluation of the Intricacies of Emotional Facial Expression of Psychiatric Patients Using Computational Models,” in *Understanding Facial Expressions in Communication*, M. K. Mandal and A. Awasthi, Eds. New Delhi: Springer India, 2015, pp. 199–226. doi: 10.1007/978-81-322-1934-7_10.
[188] K. Wang, X. Peng, J. Yang, S. Lu, and Y. Qiao, “Suppressing Uncertainties for Large-Scale Facial Expression Recognition,” *ArXiv200210392 Cs*, Mar. 2020, Accessed: Feb. 11, 2022. [Online]. Available: http://arxiv.org/abs/2002.10392
[189] Butalia, Ayesha, Maya Ingle, and Parag Kulkarni, “Facial Expression Recognition for Security,” 2012, pp. 1449–1453.
[190] Ashraf Abbas M. Al-Modwahi, Onkemetse Sebetela, Lefoko Nehemiah Batleng, and Arash Habibi Lashkari, “Facial expression recognition intelligent security system for real time surveillance,” presented at the Proc. of World Congress in Computer Science, Computer Engineering, and Applied Computing, 2012.
[191] J.-U. Garbas, T. Ruf, M. Unfried, and A. Dieckmann, “Towards Robust Real-Time Valence Recognition from Facial Expressions for Market Research Applications,” in *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction*, Geneva, Switzerland, Sep. 2013, pp. 570–575. doi: 10.1109/ACII.2013.100.
[192] G. Yolcu, I. Oztel, S. Kazan, C. Oz, and F. Bunyak, “Deep learning-based face analysis system for monitoring customer interest,” *J. Ambient Intell. Humaniz. Comput.*, vol. 11, no. 1, pp. 237–248, Jan. 2020, doi: 10.1007/s12652-019-01310-5.
[193] A. Generosi, S. Ceccacci, and M. Mengoni, “A deep learning-based system to track and analyze customer behavior in retail store,” in *2018 IEEE 8th International Conference on Consumer Electronics - Berlin (ICCE-Berlin)*, Berlin, Sep. 2018, pp. 1–6. doi: 10.1109/ICCE-Berlin.2018.8576169.
[194] Ekman, Paul and Wallace V. Friesen, *Facial action coding system*. Environmental Psychology & Nonverbal Behavior, 1978.
[195] J. Hamm, C. G. Kohler, R. C. Gur, and R. Verma, “Automated Facial Action Coding System for dynamic analysis of facial expressions in neuropsychiatric disorders,” *J.*

Neurosci. Methods, vol. 200, no. 2, pp. 237–256, 2011, doi: 10.1016/j.jneumeth.2011.06.023.
[196] S. Du, Y. Tao, and A. M. Martinez, "Compound facial expressions of emotion," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 111, no. 15, pp. E1454-1462, Apr. 2014, doi: 10.1073/pnas.1322355111.
[197] T. Kanade, J. F. Cohn, and Yingli Tian, "Comprehensive database for facial expression analysis," in *Proceedings Fourth IEEE International Conference on Automatic Face and Gesture Recognition (Cat. No. PR00580)*, Grenoble, France, 2000, pp. 46–53. doi: 10.1109/AFGR.2000.840611.
[198] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, "The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression," in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops*, San Francisco, CA, USA, Jun. 2010, pp. 94–101. doi: 10.1109/CVPRW.2010.5543262.
[199] L. Yin, X. Chen, Y. Sun, T. Worm, and M. Reale, "A high-resolution 3D dynamic facial expression database," in *2008 8th IEEE International Conference on Automatic Face & Gesture Recognition*, Amsterdam, Netherlands, Sep. 2008, pp. 1–6. doi: 10.1109/AFGR.2008.4813324.
[200] X. Zhang *et al.*, "BP4D-Spontaneous: a high-resolution spontaneous 3D dynamic facial expression database," *Image Vis. Comput.*, vol. 32, no. 10, pp. 692–706, Oct. 2014, doi: 10.1016/j.imavis.2014.06.002.
[201] Y. Ye, Z. Song, J. Guo, and Y. Qiao, "SIAT-3DFE: A High-Resolution 3D Facial Expression Dataset," *IEEE Access*, vol. 8, pp. 48205–48211, 2020, doi: 10.1109/ACCESS.2020.2979518.
[202] I. J. Goodfellow *et al.*, "Challenges in Representation Learning: A report on three machine learning contests," *ArXiv13070414 Cs Stat*, Jul. 2013, Accessed: Feb. 11, 2022. [Online]. Available: http://arxiv.org/abs/1307.0414
[203] M. Pantic, M. Valstar, R. Rademaker, and L. Maat, "Web-Based Database for Facial Expression Analysis," in *2005 IEEE International Conference on Multimedia and Expo*, Amsterdam, The Netherlands, 2005, pp. 317–321. doi: 10.1109/ICME.2005.1521424.
[204] A. Dhall, R. Goecke, S. Ghosh, J. Joshi, J. Hoey, and T. Gedeon, "From individual to group-level emotion recognition: EmotiW 5.0," in *Proceedings of the 19th ACM International Conference on Multimodal Interaction*, Glasgow UK, Nov. 2017, pp. 524–528. doi: 10.1145/3136755.3143004.
[205] S. Cheng, I. Kotsia, M. Pantic, and S. Zafeiriou, "4DFAB: A Large Scale 4D Facial Expression Database for Biometric Applications," *ArXiv171201443 Cs*, Jun. 2018, Accessed: Feb. 11, 2022. [Online]. Available: http://arxiv.org/abs/1712.01443
[206] C. F. Benitez-Quiroz, R. Srinivasan, and A. M. Martinez, "EmotioNet: An Accurate, Real-Time Algorithm for the Automatic Annotation of a Million Facial Expressions in the Wild," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 5562–5570. doi: 10.1109/CVPR.2016.600.
[207] M. Lyons, S. Akamatsu, M. Kamachi, and J. Gyoba, "Coding facial expressions with Gabor wavelets," in *Proceedings Third IEEE International Conference on Automatic Face and Gesture Recognition*, Nara, Japan, 1998, pp. 200–205. doi: 10.1109/AFGR.1998.670949.
[208] M. Taini, G. Zhao, S. Z. Li, and M. Pietikainen, "Facial expression recognition from near-infrared video sequences," in *2008 19th International Conference on Pattern Recognition*, Tampa, FL, USA, Dec. 2008, pp. 1–4. doi: 10.1109/ICPR.2008.4761697.

[209] P. Viola and M. Jones, “Rapid object detection using a boosted cascade of simple features,” in *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, Kauai, HI, USA, 2001, vol. 1, p. I-511-I–518. doi: 10.1109/CVPR.2001.990517.

[210] H. Filali, J. Riffi, A. M. Mahraz, and H. Tairi, “Multiple face detection based on machine learning,” in *2018 International Conference on Intelligent Systems and Computer Vision (ISCV)*, Fez, Apr. 2018, pp. 1–8. doi: 10.1109/ISACV.2018.8354058.

[211] M. Böhme, M. Haker, K. Riemer, T. Martinetz, and E. Barth, “Face Detection Using a Time-of-Flight Camera,” in *Dynamic 3D Imaging*, vol. 5742, A. Kolb and R. Koch, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 167–176. doi: 10.1007/978-3-642-03778-8_13.

[212] S. S. Farfade, M. Saberian, and L.-J. Li, “Multi-view Face Detection Using Deep Convolutional Neural Networks,” *ArXiv150202766 Cs*, Apr. 2015, Accessed: Feb. 11, 2022. [Online]. Available: http://arxiv.org/abs/1502.02766

[213] S. Lawrence, C. L. Giles, Ah Chung Tsoi, and A. D. Back, “Face recognition: a convolutional neural-network approach,” *IEEE Trans. Neural Netw.*, vol. 8, no. 1, pp. 98–113, Jan. 1997, doi: 10.1109/72.554195.

[214] S. L. Happy and A. Routray, “Automatic facial expression recognition using features of salient facial patches,” *IEEE Trans. Affect. Comput.*, vol. 6, no. 1, pp. 1–12, Jan. 2015, doi: 10.1109/TAFFC.2014.2386334.

[215] B. Johnston and P. de Chazal, “A review of image-based automatic facial landmark identification techniques,” *EURASIP J. Image Video Process.*, vol. 2018, no. 1, p. 86, Dec. 2018, doi: 10.1186/s13640-018-0324-4.

[216] H. Fan and E. Zhou, “Approaching human level facial landmark localization by deep learning,” *Image Vis. Comput.*, vol. 47, pp. 27–35, Mar. 2016, doi: 10.1016/j.imavis.2015.11.004.

[217] S. A. Bargal, E. Barsoum, C. C. Ferrer, and C. Zhang, “Emotion recognition in the wild from videos using images,” in *Proceedings of the 18th ACM International Conference on Multimodal Interaction*, Tokyo Japan, Oct. 2016, pp. 433–436. doi: 10.1145/2993148.2997627.

[218] D. A. Pitaloka, A. Wulandari, T. Basaruddin, and D. Y. Liliana, “Enhancing CNN with Preprocessing Stage in Automatic Emotion Recognition,” *Procedia Comput. Sci.*, vol. 116, pp. 523–529, 2017, doi: 10.1016/j.procs.2017.10.038.

[219] T. Tran, T. Pham, G. Carneiro, L. Palmer, and I. Reid, “A Bayesian Data Augmentation Approach for Learning Deep Models,” *ArXiv171010564 Cs*, Oct. 2017, Accessed: Feb. 11, 2022. [Online]. Available: http://arxiv.org/abs/1710.10564

[220] J. Lemley, S. Bazrafkan, and P. Corcoran, “Smart Augmentation - Learning an Optimal Data Augmentation Strategy,” *IEEE Access*, vol. 5, pp. 5858–5869, 2017, doi: 10.1109/ACCESS.2017.2696121.

[221] A. J. Ratner, H. R. Ehrenberg, Z. Hussain, J. Dunnmon, and C. Ré, “Learning to Compose Domain-Specific Transformations for Data Augmentation,” *ArXiv170901643 Cs Stat*, Sep. 2017, Accessed: Feb. 11, 2022. [Online]. Available: http://arxiv.org/abs/1709.01643

[222] C. Lin *et al.*, “Online Hyper-parameter Learning for Auto-Augmentation Strategy,” *ArXiv190507373 Cs*, Aug. 2019, Accessed: Feb. 11, 2022. [Online]. Available: http://arxiv.org/abs/1905.07373

[223] D. Gabor, “Theory of communication. Part 1: The analysis of information,” *J. Inst. Electr. Eng. - Part III Radio Commun. Eng.*, vol. 93, no. 26, pp. 429–441, Nov. 1946, doi: 10.1049/ji-3-2.1946.0074.

[224] T. Ojala, M. Pietikäinen, and D. Harwood, “A comparative study of texture measures with classification based on featured distributions,” *Pattern Recognit.*, vol. 29, no. 1, pp. 51–59, Jan. 1996, doi: 10.1016/0031-3203(95)00067-4.
[225] N. Dalal and B. Triggs, “Histograms of Oriented Gradients for Human Detection,” in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’05)*, San Diego, CA, USA, 2005, vol. 1, pp. 886–893. doi: 10.1109/CVPR.2005.177.
[226] A. J. Calder, A. M. Burton, P. Miller, A. W. Young, and S. Akamatsu, “A principal component analysis of facial expressions,” *Vision Res.*, vol. 41, no. 9, pp. 1179–1208, Apr. 2001, doi: 10.1016/S0042-6989(01)00002-5.
[227] D. G. Lowe, “Object recognition from local scale-invariant features,” in *Proceedings of the Seventh IEEE International Conference on Computer Vision*, Kerkyra, Greece, 1999, pp. 1150–1157 vol.2. doi: 10.1109/ICCV.1999.790410.
[228] S. Berretti, A. D. Bimbo, P. Pala, B. B. Amor, and M. Daoudi, “A Set of Selected SIFT Features for 3D Facial Expression Recognition,” in *2010 20th International Conference on Pattern Recognition*, Istanbul, Turkey, Aug. 2010, pp. 4125–4128. doi: 10.1109/ICPR.2010.1002.
[229] T. F. Cootes, G. J. Edwards, and C. J. Taylor, “Active appearance models,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 6, pp. 681–685, Jun. 2001, doi: 10.1109/34.927467.
[230] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, “Active Shape Models-Their Training and Application,” *Comput. Vis. Image Underst.*, vol. 61, no. 1, pp. 38–59, Jan. 1995, doi: 10.1006/cviu.1995.1004.
[231] L. Chen, C. Zhou, and L. Shen, “Facial Expression Recognition Based on SVM in E-learning,” *IERI Procedia*, vol. 2, pp. 781–787, 2012, doi: 10.1016/j.ieri.2012.06.171.
[232] P. Michel and R. El Kaliouby, “Real time facial expression recognition in video using support vector machines,” in *Proceedings of the 5th international conference on Multimodal interfaces - ICMI ’03*, Vancouver, British Columbia, Canada, 2003, p. 258. doi: 10.1145/958432.958479.
[233] Yubo Wang, Haizhou Ai, Bo Wu, and Chang Huang, “Real time facial expression recognition with AdaBoost,” in *Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004.*, Cambridge, UK, 2004, pp. 926-929 Vol.3. doi: 10.1109/ICPR.2004.1334680.
[234] X. Pu, K. Fan, X. Chen, L. Ji, and Z. Zhou, “Facial expression recognition from image sequences using twofold random forest classifier,” *Neurocomputing*, vol. 168, pp. 1173–1180, Nov. 2015, doi: 10.1016/j.neucom.2015.05.005.
[235] Y. Wang, Y. Li, Y. Song, and X. Rong, “Facial Expression Recognition Based on Random Forest and Convolutional Neural Network,” *Information*, vol. 10, no. 12, p. 375, Nov. 2019, doi: 10.3390/info10120375.
[236] G. Wang and J. Gong, “Facial Expression Recognition Based on Improved LeNet-5 CNN,” in *2019 Chinese Control and Decision Conference (CCDC)*, Nanchang, China, Jun. 2019, pp. 5655–5660. doi: 10.1109/CCDC.2019.8832535.
[237] X. Chen, X. Yang, M. Wang, and J. Zou, “Convolution neural network for automatic facial expression recognition,” in *2017 International Conference on Applied System Innovation (ICASI)*, Sapporo, Japan, May 2017, pp. 814–817. doi: 10.1109/ICASI.2017.7988558.
[238] Y. Chen, J. Du, Q. Liu, and B. Zeng, “Robust Expression Recognition Using ResNet with a Biologically-Plausible Activation Function,” in *Image and Video Technology*, vol. 10799, S. Satoh, Ed. Cham: Springer International Publishing, 2018, pp. 426–438. doi: 10.1007/978-3-319-92753-4_33.

[239] H. Guo and J. Chen, “Dynamic Facial Expression Recognition Based on ResNet and LSTM,” *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 790, no. 1, p. 012145, Mar. 2020, doi: 10.1088/1757-899X/790/1/012145.
[240] Z. Yu, Q. Liu, and G. Liu, “Deeper cascaded peak-piloted network for weak expression recognition,” *Vis. Comput.*, vol. 34, no. 12, pp. 1691–1699, Dec. 2018, doi: 10.1007/s00371-017-1443-0.
[241] X. Zhao *et al.*, “Peak-Piloted Deep Network for Facial Expression Recognition,” *ArXiv160706997 Cs*, Jan. 2017, Accessed: Feb. 11, 2022. [Online]. Available: http://arxiv.org/abs/1607.06997
[242] H. Jun, L. Shuai, S. Jinming, L. Yue, W. Jingwei, and J. Peng, “Facial Expression Recognition Based on VGGNet Convolutional Neural Network,” in *2018 Chinese Automation Congress (CAC)*, Xi’an, China, Nov. 2018, pp. 4146–4151. doi: 10.1109/CAC.2018.8623238.
[243] Y. Lv, Z. Feng, and C. Xu, “Facial expression recognition via deep learning,” in *2014 International Conference on Smart Computing*, Hong Kong, Hong Kong, Nov. 2014, pp. 303–308. doi: 10.1109/SMARTCOMP.2014.7043872.
[244] A. Mostafa, M. I. Khalil, and H. Abbas, “Emotion Recognition by Facial Features using Recurrent Neural Networks,” in *2018 13th International Conference on Computer Engineering and Systems (ICCES)*, Cairo, Egypt, Dec. 2018, pp. 417–422. doi: 10.1109/ICCES.2018.8639182.
[245] H. Kobayashi and F. Hara, “Dynamic recognition of basic facial expressions by discrete-time recurrent neural network,” in *Proceedings of 1993 International Conference on Neural Networks (IJCNN-93-Nagoya, Japan)*, Nagoya, Japan, 1993, vol. 1, pp. 155–158. doi: 10.1109/IJCNN.1993.713882.
[246] T. Zhang, W. Zheng, Z. Cui, Y. Zong, and Y. Li, “Spatial–Temporal Recurrent Neural Network for Emotion Recognition,” *IEEE Trans. Cybern.*, vol. 49, no. 3, pp. 839–847, Mar. 2019, doi: 10.1109/TCYB.2017.2788081.
[247] K.-E. Ko and K.-B. Sim, “Development of a Facial Emotion Recognition Method Based on Combining AAM with DBN,” in *2010 International Conference on Cyberworlds*, Singapore, Singapore, Oct. 2010, pp. 87–91. doi: 10.1109/CW.2010.65.
[248] H. Yang, Z. Zhang, and L. Yin, “Identity-Adaptive Facial Expression Recognition through Expression Regeneration Using Conditional Generative Adversarial Networks,” in *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*, Xi’an, May 2018, pp. 294–301. doi: 10.1109/FG.2018.00050.
[249] Y.-H. Lai and S.-H. Lai, “Emotion-Preserving Representation Learning via Generative Adversarial Network for Multi-View Facial Expression Recognition,” in *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*, Xi’an, May 2018, pp. 263–270. doi: 10.1109/FG.2018.00046.
[250] F. Zhang, T. Zhang, Q. Mao, and C. Xu, “Joint Pose and Expression Modeling for Facial Expression Recognition,” in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, Jun. 2018, pp. 3359–3368. doi: 10.1109/CVPR.2018.00354.
[251] J. Chen, J. Konrad, and P. Ishwar, “VGAN-Based Image Representation Learning for Privacy-Preserving Facial Expression Recognition,” *ArXiv180307100 Cs*, Sep. 2018, Accessed: Feb. 11, 2022. [Online]. Available: http://arxiv.org/abs/1803.07100
[252] Z. Meng, P. Liu, J. Cai, S. Han, and Y. Tong, “Identity-Aware Convolutional Neural Network for Facial Expression Recognition,” in *2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017)*, Washington, DC, DC, USA, May 2017, pp. 558–565. doi: 10.1109/FG.2017.140.

[253] K. Zhang, Y. Huang, Y. Du, and L. Wang, “Facial Expression Recognition Based on Deep Evolutional Spatial-Temporal Networks,” *IEEE Trans. Image Process.*, vol. 26, no. 9, pp. 4193–4203, Sep. 2017, doi: 10.1109/TIP.2017.2689999.
[254] S. L. Happy, A. George, and A. Routray, “A real time facial expression classification system using Local Binary Patterns,” in *2012 4th International Conference on Intelligent Human Computer Interaction (IHCI)*, Kharagpur, India, Dec. 2012, pp. 1–5. doi: 10.1109/IHCI.2012.6481802.
[255] D. Ghimire, S. Jeong, J. Lee, and S. H. Park, “Facial expression recognition based on local region specific features and support vector machines,” *Multimed. Tools Appl.*, vol. 76, no. 6, pp. 7803–7821, Mar. 2017, doi: 10.1007/s11042-016-3418-y.
[256] D. Ghimire and J. Lee, “Geometric Feature-Based Facial Expression Recognition in Image Sequences Using Multi-Class AdaBoost and Support Vector Machines,” *Sensors*, vol. 13, no. 6, pp. 7714–7734, Jun. 2013, doi: 10.3390/s130607714.
[257] M. Suk and B. Prabhakaran, “Real-Time Mobile Facial Expression Recognition System -- A Case Study,” in *2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops*, Columbus, OH, USA, Jun. 2014, pp. 132–137. doi: 10.1109/CVPRW.2014.25.
[258] Y. Tian, P. Luo, X. Wang, and X. Tang, “Pedestrian Detection aided by Deep Learning Semantic Tasks,” *ArXiv14120069 Cs*, Nov. 2014, Accessed: Feb. 11, 2022. [Online]. Available: http://arxiv.org/abs/1412.0069
[259] S. Deshmukh, M. Patwardhan, and A. Mahajan, “Survey on real-time facial expression recognition techniques,” *IET Biom.*, vol. 5, no. 3, pp. 155–163, Sep. 2016, doi: 10.1049/iet-bmt.2014.0104.
[260] R. Breuer and R. Kimmel, “A Deep Learning Perspective on the Origin of Facial Expressions,” *ArXiv170501842 Cs*, May 2017, Accessed: Feb. 11, 2022. [Online]. Available: http://arxiv.org/abs/1705.01842
[261] H. Jung, S. Lee, J. Yim, S. Park, and J. Kim, “Joint Fine-Tuning in Deep Neural Networks for Facial Expression Recognition,” in *2015 IEEE International Conference on Computer Vision (ICCV)*, Santiago, Chile, Dec. 2015, pp. 2983–2991. doi: 10.1109/ICCV.2015.341.
[262] K. Zhao, W.-S. Chu, and H. Zhang, “Deep Region and Multi-label Learning for Facial Action Unit Detection,” in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 3391–3399. doi: 10.1109/CVPR.2016.369.
[263] F. An and Z. Liu, “Facial expression recognition algorithm based on parameter adaptive initialization of CNN and LSTM,” *Vis. Comput.*, vol. 36, no. 3, pp. 483–498, Mar. 2020, doi: 10.1007/s00371-019-01635-4.
[264] S. Kumar, M. K. Bhuyan, and Y. Iwahori, “Multi-level uncorrelated discriminative shared Gaussian process for multi-view facial expression recognition,” *Vis. Comput.*, vol. 37, no. 1, pp. 143–159, Jan. 2021, doi: 10.1007/s00371-019-01788-2.
[265] K. Li, Y. Jin, M. W. Akram, R. Han, and J. Chen, “Facial expression recognition with convolutional neural networks via a new face cropping and rotation strategy,” *Vis. Comput.*, vol. 36, no. 2, pp. 391–404, Feb. 2020, doi: 10.1007/s00371-019-01627-4.
[266] S. Ebrahimi Kahou, V. Michalski, K. Konda, R. Memisevic, and C. Pal, “Recurrent Neural Networks for Emotion Recognition in Video,” in *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*, Seattle Washington USA, Nov. 2015, pp. 467–474. doi: 10.1145/2818346.2830596.
[267] D. H. Kim, W. J. Baddar, J. Jang, and Y. M. Ro, “Multi-Objective Based Spatio-Temporal Feature Representation Learning Robust to Expression Intensity Variations

for Facial Expression Recognition," *IEEE Trans. Affect. Comput.*, vol. 10, no. 2, pp. 223–236, Apr. 2019, doi: 10.1109/TAFFC.2017.2695999.
[268] W.-S. Chu, F. De la Torre, and J. F. Cohn, "Learning Spatial and Temporal Cues for Multi-Label Facial Action Unit Detection," in *2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017)*, Washington, DC, DC, USA, May 2017, pp. 25–32. doi: 10.1109/FG.2017.13.
[269] B. Hasani and M. H. Mahoor, "Facial Expression Recognition Using Enhanced Deep 3D Convolutional Neural Networks," *2017 IEEE Conf. Comput. Vis. Pattern Recognit. Workshop CVPRW*, pp. 2278–2288, Jul. 2017, doi: 10.1109/CVPRW.2017.282.
[270] F. Nonis, N. Dagnes, F. Marcolin, and E. Vezzetti, "3D Approaches and Challenges in Facial Expression Recognition Algorithms—A Literature Review," *Appl. Sci.*, vol. 9, no. 18, p. 3904, Sep. 2019, doi: 10.3390/app9183904.
[271] A. Azazi, S. Lebai Lutfi, I. Venkat, and F. Fernández-Martínez, "Towards a robust affect recognition: Automatic facial expression recognition in 3D faces," *Expert Syst. Appl.*, vol. 42, no. 6, pp. 3056–3066, Apr. 2015, doi: 10.1016/j.eswa.2014.10.042.
[272] Xudong Yang, Di Huang, Yunhong Wang, and Liming Chen, "Automatic 3D facial expression recognition using geometric scattering representation," in *2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*, Ljubljana, May 2015, pp. 1–6. doi: 10.1109/FG.2015.7163090.
[273] A. Jan, H. Ding, H. Meng, L. Chen, and H. Li, "Accurate Facial Parts Localization and Deep Learning for 3D Facial Expression Recognition," *ArXiv180305846 Cs*, Mar. 2018, Accessed: Feb. 11, 2022. [Online]. Available: http://arxiv.org/abs/1803.05846
[274] V. Le, H. Tang, and T. S. Huang, "Expression recognition from 3D dynamic faces using robust spatio-temporal shape features," in *Face and Gesture 2011*, Santa Barbara, CA, USA, Mar. 2011, pp. 414–421. doi: 10.1109/FG.2011.5771435.
[275] G. Sandbach, S. Zafeiriou, M. Pantic, and D. Rueckert, "A dynamic approach to the recognition of 3D facial expressions and their temporal models," in *Face and Gesture 2011*, Santa Barbara, CA, USA, Mar. 2011, pp. 406–413. doi: 10.1109/FG.2011.5771434.
[276] G. Sandbach, S. Zafeiriou, M. Pantic, and D. Rueckert, "Recognition of 3D facial expression dynamics," *Image Vis. Comput.*, vol. 30, no. 10, pp. 762–773, Oct. 2012, doi: 10.1016/j.imavis.2012.01.006.
[277] X.-P. Huynh, T.-D. Tran, and Y.-G. Kim, "Convolutional Neural Network Models for Facial Expression Recognition Using BU-3DFE Database," in *Information Science and Applications (ICISA) 2016*, vol. 376, K. J. Kim and N. Joukov, Eds. Singapore: Springer Singapore, 2016, pp. 441–450. doi: 10.1007/978-981-10-0557-2_44.
[278] W. Zeng, H. Li, L. Chen, J.-M. Morvan, and X. D. Gu, "An automatic 3D expression recognition framework based on sparse representation of conformal images," in *2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*, Shanghai, China, Apr. 2013, pp. 1–8. doi: 10.1109/FG.2013.6553749.
[279] Hassen Drira, M. Daoudi, Anuj Srivastava, and Stefano Berretti, "3D dynamic expression recognition based on a novel Deformation Vector Field and Random Forest," presented at the International Conference on Pattern Recognition, 2012.
[280] Mohamed Daoudi, Hassen Drira, Boulbaba Ben Amor, and Stefano Berretti, "A dynamic geometry-based approach for 4D facial expressions recognition," presented at the Visual Information Processing, 2013.
[281] X. Zhao, D. Huang, E. Dellandrea, and L. Chen, "Automatic 3D Facial Expression Recognition Based on a Bayesian Belief Net and a Statistical Facial Feature Model," in

2010 20th International Conference on Pattern Recognition, Istanbul, Turkey, Aug. 2010, pp. 3724–3727. doi: 10.1109/ICPR.2010.907.
[282] Hui Chen, Jiangdong Li, Fengjun Zhang, Yang Li, and Hongan Wang, "3D model-based continuous emotion recognition," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, USA, Jun. 2015, pp. 1836–1845. doi: 10.1109/CVPR.2015.7298793.
[283] D. Fabiano and S. Canavan, "Spontaneous and Non-Spontaneous 3D Facial Expression Recognition Using a Statistical Model with Global and Local Constraints," in *2018 25th IEEE International Conference on Image Processing (ICIP)*, Athens, Oct. 2018, pp. 3089–3093. doi: 10.1109/ICIP.2018.8451171.
[284] O. Ocegueda, T. Fang, S. K. Shah, and I. A. Kakadiaris, "Expressive Maps for 3D Facial Expression Recognition," in *2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*, Barcelona, Spain, Nov. 2011, pp. 1270–1275. doi: 10.1109/ICCVW.2011.6130397.
[285] A. Savran and B. Sankur, "Non-rigid registration-based model-free 3D facial expression recognition," *Comput. Vis. Image Underst.*, vol. 162, pp. 146–165, Sep. 2017, doi: 10.1016/j.cviu.2017.07.005.
[286] A. Savran and B. Sankur, "Non-rigid registration of 3D surfaces by deformable 2D triangular meshes," in *2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, Anchorage, AK, USA, Jun. 2008, pp. 1–6. doi: 10.1109/CVPRW.2008.4563083.
[287] A. Savran, B. Sankur, and M. Taha Bilge, "Regression-based intensity estimation of facial action units," *Image Vis. Comput.*, vol. 30, no. 10, pp. 774–784, Oct. 2012, doi: 10.1016/j.imavis.2011.11.008.
[288] A. Savran, B. Sankur, and M. Taha Bilge, "Comparative evaluation of 3D vs. 2D modality for automatic detection of facial action units," *Pattern Recognit.*, vol. 45, no. 2, pp. 767–782, Feb. 2012, doi: 10.1016/j.patcog.2011.07.022.
[289] N. Alyuz, B. Gokberk, H. Dibeklioglu, and L. Akarun, "Component-based registration with curvature descriptors for expression insensitive 3d face recognition," in *2008 8th IEEE International Conference on Automatic Face & Gesture Recognition*, Amsterdam, Sep. 2008, pp. 1–6. doi: 10.1109/AFGR.2008.4813359.
[290] S. Berretti, B. Ben Amor, M. Daoudi, and A. del Bimbo, "3D facial expression recognition using SIFT descriptors of automatically detected keypoints," *Vis. Comput.*, vol. 27, no. 11, pp. 1021–1036, Nov. 2011, doi: 10.1007/s00371-011-0611-x.
[291] S. Berretti, A. del Bimbo, and P. Pala, "Automatic facial expression recognition in real-time from dynamic sequences of 3D face scans," *Vis. Comput.*, vol. 29, no. 12, pp. 1333–1350, Dec. 2013, doi: 10.1007/s00371-013-0869-2.
[292] P. Zarbakhsh and H. Demirel, "4D facial expression recognition using multimodal time series analysis of geometric landmark-based deformations," *Vis. Comput.*, vol. 36, no. 5, pp. 951–965, May 2020, doi: 10.1007/s00371-019-01705-7.
[293] P. Lemaire, M. Ardabilian, L. Chen, and M. Daoudi, "Fully automatic 3D facial expression recognition using differential mean curvature maps and histograms of oriented gradients," in *2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*, Shanghai, China, Apr. 2013, pp. 1–7. doi: 10.1109/FG.2013.6553821.
[294] C. A. Corneanu, M. O. Simon, J. F. Cohn, and S. E. Guerrero, "Survey on RGB, 3D, Thermal, and Multimodal Approaches for Facial Expression Recognition: History, Trends, and Affect-Related Applications," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 8, pp. 1548–1568, Aug. 2016, doi: 10.1109/TPAMI.2016.2515606.

[295] M. Hayat and M. Bennamoun, “An Automatic Framework for Textured 3D Video-Based Facial Expression Recognition,” *IEEE Trans. Affect. Comput.*, vol. 5, no. 3, pp. 301–313, Jul. 2014, doi: 10.1109/TAFFC.2014.2330580.
[296] A. Jan and Hongying Meng, “Automatic 3D facial expression recognition using geometric and textured feature fusion,” in *2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*, Ljubljana, May 2015, pp. 1–6. doi: 10.1109/FG.2015.7284860.
[297] O. K. Oyedotun, G. Demisse, A. E. R. Shabayek, D. Aouada, and B. Ottersten, “Facial Expression Recognition via Joint Deep Learning of RGB-Depth Map Latent Representations,” in *2017 IEEE International Conference on Computer Vision Workshops (ICCVW)*, Venice, Italy, Oct. 2017, pp. 3161–3168. doi: 10.1109/ICCVW.2017.374.
[298] H. Yang and L. Yin, “CNN based 3D facial expression recognition using masking and landmark features,” in *2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII)*, San Antonio, TX, Oct. 2017, pp. 556–560. doi: 10.1109/ACII.2017.8273654.
[299] H. Li, J. Sun, D. Wang, Z. Xu, and L. Chen, “Deep Representation of Facial Geometric and Photometric Attributes for Automatic 3D Facial Expression Recognition,” *ArXiv151103015 Cs*, Nov. 2015, Accessed: Feb. 11, 2022. [Online]. Available: http://arxiv.org/abs/1511.03015
[300] Z. Chen, D. Huang, Y. Wang, and L. Chen, “Fast and Light Manifold CNN based 3D Facial Expression Recognition across Pose Variations,” in *Proceedings of the 26th ACM international conference on Multimedia*, Seoul Republic of Korea, Oct. 2018, pp. 229–238. doi: 10.1145/3240508.3240568.
[301] Y. Guo, H. Wang, Q. Hu, H. Liu, L. Liu, and M. Bennamoun, “Deep Learning for 3D Point Clouds: A Survey,” *ArXiv191212033 Cs Eess*, Jun. 2020, Accessed: Feb. 11, 2022. [Online]. Available: http://arxiv.org/abs/1912.12033
[302] C. R. Qi, H. Su, K. Mo, and L. J. Guibas, “PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation,” *ArXiv161200593 Cs*, Apr. 2017, Accessed: Feb. 11, 2022. [Online]. Available: http://arxiv.org/abs/1612.00593
[303] C. R. Qi, L. Yi, H. Su, and L. J. Guibas, “PointNet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space,” *ArXiv170602413 Cs*, Jun. 2017, Accessed: Feb. 11, 2022. [Online]. Available: http://arxiv.org/abs/1706.02413
[304] K. Y. Choi, “Analysis of Facial Asymmetry,” *Arch. Craniofacial Surg.*, vol. 16, no. 1, p. 1, 2015, doi: 10.7181/acfs.2015.16.1.1.
[305] W. Wei, E. S. L. Ho, K. D. McCay, R. Damaševičius, R. Maskeliūnas, and A. Esposito, “Assessing Facial Symmetry and Attractiveness using Augmented Reality,” *Pattern Anal. Appl.*, Mar. 2021, doi: 10.1007/s10044-021-00975-z.
[306] A. C. Little, B. C. Jones, and L. M. DeBruine, “Facial attractiveness: evolutionary based research,” *Philos. Trans. R. Soc. B Biol. Sci.*, vol. 366, no. 1571, pp. 1638–1659, Jun. 2011, doi: 10.1098/rstb.2010.0404.
[307] R. E. Riggio and S. B. Woll, “The Role of Nonverbal Cues and Physical Attractiveness in the Selection of Dating Partners,” *J. Soc. Pers. Relatsh.*, vol. 1, no. 3, pp. 347–357, Sep. 1984, doi: 10.1177/0265407584013007.
[308] S. N. Talamas, K. I. Mavor, and D. I. Perrett, “Blinded by Beauty: Attractiveness Bias and Accurate Perceptions of Academic Performance,” *PLOS ONE*, vol. 11, no. 2, p. e0148284, Feb. 2016, doi: 10.1371/journal.pone.0148284.
[309] R. R. Jurin, D. Roush, and J. Danter, “Communicating Without Words,” in *Environmental Communication. Second Edition*, Dordrecht: Springer Netherlands, 2010, pp. 221–230. doi: 10.1007/978-90-481-3987-3_14.

[310] G. Croxson, M. May, and S. J. Mester, “Grading facial nerve function: House-Brackmann versus Burres-Fisch methods,” *Am. J. Otol.*, vol. 11, no. 4, pp. 240–246, Jul. 1990.
[311] “Rapid Grading System (RGS).” [Online]. Available: http://www.entusa.com/bells_palsy.htm
[312] A. Ahrens, D. Skarada, M. Wallace, J. Y. Cheung, and J. G. Neely, “Rapid simultaneous comparison system for subjective grading scales grading scales for facial paralysis,” *Am. J. Otol.*, vol. 20, no. 5, pp. 667–671, Sep. 1999.
[313] A. A. Pourmomeny, H. Zadmehr, and M. Hossaini, “Measurement of facial movements with Photoshop software during treatment of facial nerve palsy,” *J. Res. Med. Sci. Off. J. Isfahan Univ. Med. Sci.*, vol. 16, no. 10, pp. 1313–1318, Oct. 2011.
[314] L. Sjögreen, A. Lohmander, and S. Kiliaridis, “Exploring quantitative methods for evaluation of lip function,” *J. Oral Rehabil.*, vol. 38, no. 6, pp. 410–422, Jun. 2011, doi: 10.1111/j.1365-2842.2010.02168.x.
[315] A. Gaber, M. F. Faher, and M. A. Waned, “Automated grading of facial paralysis using the Kinect v2: A proof of concept study,” in *2015 International Conference on Virtual Rehabilitation (ICVR)*, Valencia, Spain, Jun. 2015, pp. 258–264. doi: 10.1109/ICVR.2015.7358577.
[316] X.-F. Han, J. S. Jin, M.-J. Wang, W. Jiang, L. Gao, and L. Xiao, “A review of algorithms for filtering the 3D point cloud,” *Signal Process. Image Commun.*, vol. 57, pp. 103–112, Sep. 2017, doi: 10.1016/j.image.2017.05.009.
[317] R. B. Rusu and S. Cousins, “3D is here: Point Cloud Library (PCL),” in *2011 IEEE International Conference on Robotics and Automation*, Shanghai, China, May 2011, pp. 1–4. doi: 10.1109/ICRA.2011.5980567.
[318] R. Rostami, F. S. Bashiri, B. Rostami, and Z. Yu, “A Survey on Data-Driven 3D Shape Descriptors,” *Comput. Graph. Forum*, vol. 38, no. 1, pp. 356–393, Feb. 2019, doi: 10.1111/cgf.13536.
[319] R. Osada, T. Funkhouser, B. Chazelle, and D. Dobkin, “Matching 3D models with shape distributions,” in *Proceedings International Conference on Shape Modeling and Applications*, Genova, Italy, 2001, pp. 154–166. doi: 10.1109/SMA.2001.923386.
[320] M. Novotni and R. Klein, “Shape retrieval using 3D Zernike descriptors,” *Comput.-Aided Des.*, vol. 36, no. 11, pp. 1047–1062, Sep. 2004, doi: 10.1016/j.cad.2004.01.005.
[321] D. Saupe and D. V. Vranić, “3D Model Retrieval with Spherical Harmonics and Moments,” in *Pattern Recognition*, vol. 2191, B. Radig and S. Florczyk, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2001, pp. 392–397. doi: 10.1007/3-540-45404-7_52.
[322] V. Jain and H. Zhang, “A spectral approach to shape-based retrieval of articulated 3D models,” *Comput.-Aided Des.*, vol. 39, no. 5, pp. 398–407, May 2007, doi: 10.1016/j.cad.2007.02.009.
[323] J. Assfalg, M. Bertini, A. D. Bimbo, and P. Pala, “Content-Based Retrieval of 3-D Objects Using Spin Image Signatures,” *IEEE Trans. Multimed.*, vol. 9, no. 3, pp. 589–599, Apr. 2007, doi: 10.1109/TMM.2006.886271.
[324] A. Zaharescu, E. Boyer, K. Varanasi, and R. Horaud, “Surface feature detection and description with applications to mesh matching,” in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, Miami, FL, Jun. 2009, pp. 373–380. doi: 10.1109/CVPR.2009.5206748.
[325] Marcel Kortgen, Gil-Joo Park, Marcin Novotni, and Reinhard Klein, “3D shape matching with 3D shape contexts,” presented at the Central European Seminar on Computer Graphics, Jan. 2003.

[326] X.-F. Han, S.-J. Sun, X.-Y. Song, and G.-Q. Xiao, “3D Point Cloud Descriptors in Hand-crafted and Deep Learning Age: State-of-the-Art,” *ArXiv180202297 Cs*, Jul. 2020, Accessed: Feb. 16, 2022. [Online]. Available: http://arxiv.org/abs/1802.02297
[327] B. Kolb and R. Gibb, “Brain plasticity and behaviour in the developing brain,” *J. Can. Acad. Child Adolesc. Psychiatry J. Acad. Can. Psychiatr. Enfant Adolesc.*, vol. 20, no. 4, pp. 265–276, Nov. 2011.
[328] A. Erdemir, S. McLean, W. Herzog, and A. J. van den Bogert, “Model-based estimation of muscle forces exerted during movements,” *Clin. Biomech. Bristol Avon*, vol. 22, no. 2, pp. 131–154, Feb. 2007, doi: 10.1016/j.clinbiomech.2006.09.005.
[329] S. He, J. J. Soraghan, B. F. O’Reilly, and D. Xing, “Quantitative Analysis of Facial Paralysis Using Local Binary Patterns in Biomedical Videos,” *IEEE Trans. Biomed. Eng.*, vol. 56, no. 7, pp. 1864–1870, Jul. 2009, doi: 10.1109/TBME.2009.2017508.
[330] D. Jayatilake, T. Isezaki, Y. Teramoto, K. Eguchi, and K. Suzuki, “Robot Assisted Physiotherapy to Support Rehabilitation of Facial Paralysis,” *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 22, no. 3, pp. 644–653, May 2014, doi: 10.1109/TNSRE.2013.2279169.
[331] M. Eskes, A. J. M. Balm, M. J. A. van Alphen, L. E. Smeele, I. Stavness, and F. van der Heijden, “Simulation of facial expressions using person-specific sEMG signals controlling a biomechanical face model,” *Int. J. Comput. Assist. Radiol. Surg.*, vol. 13, no. 1, pp. 47–59, Jan. 2018, doi: 10.1007/s11548-017-1659-5.
[332] G. Pileicikiene, E. Varpiotas, R. Surna, and A. Surna, “A three-dimensional model of the human masticatory system, including the mandible, the dentition and the temporomandibular joints,” *Stomatologija*, vol. 9, no. 1, pp. 27–32, 2007.
[333] Y. Zhang, E. C. Prakash, and E. Sung, “Face alive,” *J. Vis. Lang. Comput.*, vol. 15, no. 2, pp. 125–160, Apr. 2004, doi: 10.1016/j.jvlc.2003.11.002.
[334] P. Claes, D. Vandermeulen, S. De Greef, G. Willems, J. G. Clement, and P. Suetens, “Computerized craniofacial reconstruction: Conceptual framework and review,” *Forensic Sci. Int.*, vol. 201, no. 1–3, pp. 138–145, Sep. 2010, doi: 10.1016/j.forsciint.2010.03.008.
[335] W. Mollemans, F. Schutyser, N. Nadjmi, F. Maes, and P. Suetens, “Predicting soft tissue deformations for a maxillofacial surgery planning system: From computational strategies to a complete clinical validation,” *Med. Image Anal.*, vol. 11, no. 3, pp. 282–301, Jun. 2007, doi: 10.1016/j.media.2007.02.003.
[336] H. Kim, P. Jürgens, S. Weber, L.-P. Nolte, and M. Reyes, “A new soft-tissue simulation strategy for cranio-maxillofacial surgery using facial muscle template model,” *Prog. Biophys. Mol. Biol.*, vol. 103, no. 2–3, pp. 284–291, Dec. 2010, doi: 10.1016/j.pbiomolbio.2010.09.004.
[337] A. G. Hannam, “Current computational modelling trends in craniomandibular biomechanics and their clinical implications,” *J. Oral Rehabil.*, vol. 38, no. 3, pp. 217–234, Mar. 2011, doi: 10.1111/j.1365-2842.2010.02149.x.
[338] Ang-Xiao FAN, “Geometric and numerical modeling of facial mimics derived from Magnetic Resonance Imaging (MRI) using Finite Element Method,”, Ph.D. dissertation, Université de Technologie de Compiègne, 2016.
[339] J. T. Dennerlein, “Finger flexor tendon forces are a complex function of finger joint motions and fingertip forces,” *J. Hand Ther. Off. J. Am. Soc. Hand Ther.*, vol. 18, no. 2, pp. 120–127, Jun. 2005, doi: 10.1197/j.jht.2005.01.011.
[340] E. Otten, “Inverse and forward dynamics: models of multi-body systems,” *Philos. Trans. R. Soc. Lond. B. Biol. Sci.*, vol. 358, no. 1437, pp. 1493–1500, Sep. 2003, doi: 10.1098/rstb.2003.1354.

[341] O. Gottesman *et al.*, “Guidelines for reinforcement learning in healthcare,” *Nat. Med.*, vol. 25, no. 1, pp. 16–18, Jan. 2019, doi: 10.1038/s41591-018-0310-5.
[342] P. Kormushev, S. Calinon, and D. Caldwell, “Reinforcement Learning in Robotics: Applications and Real-World Challenges,” *Robotics*, vol. 2, no. 3, pp. 122–148, Jul. 2013, doi: 10.3390/robotics2030122.
[343] I. Szita, “Reinforcement Learning in Games,” in *Reinforcement Learning*, vol. 12, M. Wiering and M. van Otterlo, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 539–577. doi: 10.1007/978-3-642-27645-3_17.
[344] A. E. Sallab, M. Abdou, E. Perot, and S. Yogamani, “Deep Reinforcement Learning framework for Autonomous Driving,” *Electron. Imaging*, vol. 2017, no. 19, pp. 70–76, Jan. 2017, doi: 10.2352/ISSN.2470-1173.2017.19.AVM-023.
[345] M. Bellver, X. Giro-i-Nieto, F. Marques, and J. Torres, “Hierarchical Object Detection with Deep Reinforcement Learning,” *ArXiv161103718 Cs*, Nov. 2016, Accessed: Feb. 16, 2022. [Online]. Available: http://arxiv.org/abs/1611.03718
[346] A. Jonsson, “Deep Reinforcement Learning in Medicine,” *Kidney Dis. Basel Switz.*, vol. 5, no. 1, pp. 18–22, Feb. 2019, doi: 10.1159/000492670.
[347] T. V. Maia and M. J. Frank, “From reinforcement learning models to psychiatric and neurological disorders,” *Nat. Neurosci.*, vol. 14, no. 2, pp. 154–162, Feb. 2011, doi: 10.1038/nn.2723.
[348] K. Nowakowski *et al.*, “Human locomotion with reinforcement learning using bioinspired reward reshaping strategies,” *Med. Biol. Eng. Comput.*, vol. 59, no. 1, pp. 243–256, Jan. 2021, doi: 10.1007/s11517-020-02309-3.
[349] J. Izawa, T. Kondo, and K. Ito, “Biological arm motion through reinforcement learning,” *Biol. Cybern.*, vol. 91, no. 1, pp. 10–22, Jul. 2004, doi: 10.1007/s00422-004-0485-3.
[350] Alex Broad, “Generating muscle driven arm movements using reinforcement learning,” 2011.
[351] A. H. Abdi, P. Saha, P. Srungarapu, and S. Fels, “Muscle Excitation Estimation in Biomechanical Simulation Using NAF Reinforcement Learning,” *ArXiv180906121 Cs Stat*, pp. 133–141, 2020, doi: 10.1007/978-3-030-15923-8_11.
[352] Joshua D. Rosenberg, “Facial Nerve Paralysis Photo Gallery.” [Online]. Available: https://www.drjoshuarosenberg.com/facial-nerve-paralysis-photo-gallery/
[353] C. Sagonas, E. Antonakos, G. Tzimiropoulos, S. Zafeiriou, and M. Pantic, “300 Faces In-The-Wild Challenge: database and results,” *Image Vis. Comput.*, vol. 47, pp. 3–18, Mar. 2016, doi: 10.1016/j.imavis.2016.01.002.
[354] P. N. Belhumeur, D. W. Jacobs, D. J. Kriegman, and N. Kumar, “Localizing Parts of Faces Using a Consensus of Exemplars,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 12, pp. 2930–2940, Dec. 2013, doi: 10.1109/TPAMI.2013.23.
[355] I. Matthews and S. Baker, “Active Appearance Models Revisited,” *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 135–164, Nov. 2004, doi: 10.1023/B:VISI.0000029666.37597.d3.
[356] D. F. Dementhon and L. S. Davis, “Model-based object pose in 25 lines of code,” *Int. J. Comput. Vis.*, vol. 15, no. 1–2, pp. 123–141, Jun. 1995, doi: 10.1007/BF01450852.
[357] Y. Feng, H. Feng, M. J. Black, and T. Bolkart, “Learning an Animatable Detailed 3D Face Model from In-The-Wild Images,” *ArXiv201204012 Cs*, Jun. 2021, Accessed: Jan. 13, 2022. [Online]. Available: http://arxiv.org/abs/2012.04012
[358] T. Li, T. Bolkart, M. J. Black, H. Li, and J. Romero, “Learning a model of facial shape and expression from 4D scans,” *ACM Trans. Graph.*, vol. 36, no. 6, pp. 1–17, Nov. 2017, doi: 10.1145/3130800.3130813.

[359] R. Ramamoorthi and P. Hanrahan, “An efficient representation for irradiance environment maps,” in *Proceedings of the 28th annual conference on Computer graphics and interactive techniques - SIGGRAPH ’01*, Not Known, 2001, pp. 497–500. doi: 10.1145/383259.383317.

[360] Y. Wang, X. Tao, X. Qi, X. Shen, and J. Jia, “Image Inpainting via Generative Multi-column Convolutional Neural Networks,” *ArXiv181008771 Cs*, Oct. 2018, Accessed: Jan. 13, 2022. [Online]. Available: http://arxiv.org/abs/1810.08771

[361] Y. Deng, J. Yang, S. Xu, D. Chen, Y. Jia, and X. Tong, “Accurate 3D Face Reconstruction with Weakly-Supervised Learning: From Single Image to Image Set,” *ArXiv190308527 Cs*, Apr. 2020, Accessed: Jan. 13, 2022. [Online]. Available: http://arxiv.org/abs/1903.08527

[362] A. Bulat and G. Tzimiropoulos, “How far are we from solving the 2D & 3D Face Alignment problem? (and a dataset of 230,000 3D facial landmarks),” *2017 IEEE Int. Conf. Comput. Vis. ICCV*, pp. 1021–1030, Oct. 2017, doi: 10.1109/ICCV.2017.116.

[363] N. Aspert, D. Santa-Cruz, and T. Ebrahimi, “MESH: measuring errors between surfaces using the Hausdorff distance,” in *Proceedings. IEEE International Conference on Multimedia and Expo*, Lausanne, Switzerland, 2002, pp. 705–708. doi: 10.1109/ICME.2002.1035879.

[364] E. Karp, E. Waselchuk, C. Landis, J. Fahnhorst, B. Lindgren, and S. Lyford-Pike, “Facial Rehabilitation as Noninvasive Treatment for Chronic Facial Nerve Paralysis,” *Otol. Neurotol.*, vol. 40, no. 2, pp. 241–245, Feb. 2019, doi: 10.1097/MAO.0000000000002107.

[365] J. Hamm, C. G. Kohler, R. C. Gur, and R. Verma, “Automated Facial Action Coding System for dynamic analysis of facial expressions in neuropsychiatric disorders,” *J. Neurosci. Methods*, vol. 200, no. 2, pp. 237–256, Sep. 2011, doi: 10.1016/j.jneumeth.2011.06.023.

[366] Z. Pan *et al.*, “Clinical application of an automatic facial recognition system based on deep learning for diagnosis of Turner syndrome,” *Endocrine*, vol. 72, no. 3, pp. 865–873, Jun. 2021, doi: 10.1007/s12020-020-02539-3.

[367] D. Wu *et al.*, “Facial Recognition Intensity in Disease Diagnosis Using Automatic Facial Recognition,” *J. Pers. Med.*, vol. 11, no. 11, p. 1172, Nov. 2021, doi: 10.3390/jpm11111172.

[368] P. Urbanová, Z. Ferková, M. Jandová, M. Jurda, D. Černý, and J. Sochor, “Introducing the FIDENTIS 3D Face Database,” *Anthropol. Rev.*, vol. 81, no. 2, pp. 202–223, Jun. 2018, doi: 10.2478/anre-2018-0016.

[369] A. Ranjan, T. Bolkart, S. Sanyal, and M. J. Black, “Generating 3D faces using Convolutional Mesh Autoencoders,” *ArXiv180710267 Cs*, Jul. 2018, Accessed: Jan. 13, 2022. [Online]. Available: http://arxiv.org/abs/1807.10267

[370] Z.-H. Jiang, Q. Wu, K. Chen, and J. Zhang, “Disentangled Representation Learning for 3D Face Shape,” *ArXiv190209887 Cs*, Mar. 2019, Accessed: Jan. 13, 2022. [Online]. Available: http://arxiv.org/abs/1902.09887

[371] Radu Bogdan Rusu, N. Blodow, Z. Marton, A. Soos, and M. Beetz, “Towards 3D object maps for autonomous household robots,” in *2007 IEEE/RSJ International Conference on Intelligent Robots and Systems*, San Diego, CA, USA, Oct. 2007, pp. 3191–3198. doi: 10.1109/IROS.2007.4399309.

[372] N. S. Keskar, D. Mudigere, J. Nocedal, M. Smelyanskiy, and P. T. P. Tang, “On Large-Batch Training for Deep Learning: Generalization Gap and Sharp Minima,” *ArXiv160904836 Cs Math*, Feb. 2017, Accessed: Feb. 11, 2022. [Online]. Available: http://arxiv.org/abs/1609.04836

[373] N. Vretos, N. Nikolaidis, and I. Pitas, “3D facial expression recognition using Zernike moments on depth images,” in *2011 18th IEEE International Conference on Image Processing*, Brussels, Belgium, Sep. 2011, pp. 773–776. doi: 10.1109/ICIP.2011.6116669.
[374] Y. Wang, M. Meng, and Q. Zhen, “Learning Encoded Facial Curvature Information for 3D Facial Emotion Recognition,” in *2013 Seventh International Conference on Image and Graphics*, Qingdao, China, Jul. 2013, pp. 529–532. doi: 10.1109/ICIG.2013.112.
[375] W. Hariri, H. Tabia, N. Farah, A. Benouareth, and D. Declercq, “3D facial expression recognition using kernel methods on Riemannian manifold,” *Eng. Appl. Artif. Intell.*, vol. 64, pp. 25–32, Sep. 2017, doi: 10.1016/j.engappai.2017.05.009.
[376] S. Wallace, M. Coleman, and A. Bailey, “An investigation of basic facial expression recognition in autism spectrum disorders,” *Cogn. Emot.*, vol. 22, no. 7, pp. 1353–1380, Nov. 2008, doi: 10.1080/02699930701782153.
[377] G. Muhammad, M. Alsulaiman, S. U. Amin, A. Ghoneim, and M. F. Alhamid, “A Facial-Expression Monitoring System for Improved Healthcare in Smart Cities,” *IEEE Access*, vol. 5, pp. 10871–10881, 2017, doi: 10.1109/ACCESS.2017.2712788.
[378] M. M. Bronstein, J. Bruna, Y. LeCun, A. Szlam, and P. Vandergheynst, “Geometric deep learning: going beyond Euclidean data,” *IEEE Signal Process. Mag.*, vol. 34, no. 4, pp. 18–42, Jul. 2017, doi: 10.1109/MSP.2017.2693418.
[379] S. Greengard, “Geometric deep learning advances data science,” *Commun. ACM*, vol. 64, no. 1, pp. 13–15, Jan. 2021, doi: 10.1145/3433951.
[380] Charlie D. Frowd, Bogdan Matuszewski, Lik Shark, and Wei Quan, “Towards a comprehensive 3D dynamic facial expression database,” 2009, pp. 113–119.
[381] B. J. Matuszewski, W. Quan, and L.-K. Shark, “High-resolution comprehensive 3-D dynamic database for facial articulation analysis,” in *2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*, Barcelona, Spain, Nov. 2011, pp. 2128–2135. doi: 10.1109/ICCVW.2011.6130511.
[382] Farnsworth, B, “Facial Action Coding System (FACS)—A Visual Guidebook,” 2018. [Online]. Available: https://imotions. com/blog/facial-action-coding-system/
[383] S. P. Singh, L. Wang, S. Gupta, H. Goli, P. Padmanabhan, and B. Gulyás, “3D Deep Learning on Medical Images: A Review,” *ArXiv200400218 Cs Eess Q-Bio*, Oct. 2020, Accessed: Feb. 11, 2022. [Online]. Available: http://arxiv.org/abs/2004.00218
[384] T. T. Dao, “From deep learning to transfer learning for the prediction of skeletal muscle forces,” *Med. Biol. Eng. Comput.*, vol. 57, no. 5, pp. 1049–1058, May 2019, doi: 10.1007/s11517-018-1940-y.
[385] E. Alexiou, I. Viola, and P. Cesar, “PointPCA: Point Cloud Objective Quality Assessment Using PCA-Based Descriptors,” *ArXiv211112663 Cs*, Nov. 2021, Accessed: Feb. 16, 2022. [Online]. Available: http://arxiv.org/abs/2111.12663
[386] M. A. Nazari, P. Perrier, M. Chabanas, and Y. Payan, “Simulation of dynamic orofacial movements using a constitutive law varying with muscle activation,” *Comput. Methods Biomech. Biomed. Engin.*, vol. 13, no. 4, pp. 469–482, Aug. 2010, doi: 10.1080/10255840903505147.
[387] M. Bucki, M. A. Nazari, and Y. Payan, “Finite element speaker-specific face model generation for the study of speech production,” *Comput. Methods Biomech. Biomed. Engin.*, vol. 13, no. 4, pp. 459–467, Aug. 2010, doi: 10.1080/10255840903505139.
[388] J. E. Lloyd, I. Stavness, and S. Fels, “ArtiSynth: A Fast Interactive Biomechanical Modeling Toolkit Combining Multibody and Finite Element Simulation,” in *Soft Tissue Biomechanical Modeling for Computer Assisted Surgery*, vol. 11, Y. Payan, Ed. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 355–394. doi: 10.1007/8415_2012_126.

[389] I. Stavness *et al.*, “Coupled Biomechanical Modeling of the Face, Jaw, Skull, Tongue, and Hyoid Bone,” in *3D Multiscale Physiological Human*, N. Magnenat-Thalmann, O. Ratib, and H. F. Choi, Eds. London: Springer London, 2014, pp. 253–274. doi: 10.1007/978-1-4471-6275-9_11.

[390] C. Flynn, M. A. Nazari, P. Perrier, S. Fels, P. M. F. Nielsen, and Y. Payan, “Computational Modeling of the Passive and Active Components of the Face,” in *Biomechanics of Living Organs*, Elsevier, 2017, pp. 377–394. doi: 10.1016/B978-0-12-804009-6.00018-3.

[391] A. H. Abdi, M. Malakoutian, T. Oxland, and S. Fels, “Reinforcement Learning for High-dimensional Continuous Control in Biomechanics: An Intro to ArtiSynth-RL,” *ArXiv191013859 Eess*, Dec. 2019, Accessed: Feb. 16, 2022. [Online]. Available: http://arxiv.org/abs/1910.13859

[392] Y.-I. Tian, T. Kanade, and J. F. Cohn, “Recognizing action units for facial expression analysis,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 2, pp. 97–115, Feb. 2001, doi: 10.1109/34.908962.

[393] S. Song *et al.*, “Deep reinforcement learning for modeling human locomotion control in neuromechanical simulation,” *J. NeuroEngineering Rehabil.*, vol. 18, no. 1, p. 126, Dec. 2021, doi: 10.1186/s12984-021-00919-y.

[394] H. Anderl, “Reconstruction of the face through cross-face-nerve transplantation in facial paralysis,” *Chir. Plast.*, vol. 2, no. 1, pp. 17–45, 1973, doi: 10.1007/BF00280913.

[395] J. Lopez, E. D. Rodriguez, and A. H. Dorafshar, *Facial Transplantation*. Elsevier Inc., 2019. doi: 10.1016/B978-0-323-49755-8.00051-7.

[396] M. W. Robinson and J. Baiungo, “Facial Rehabilitation: Evaluation and Treatment Strategies for the Patient with Facial Palsy,” *Otolaryngol. Clin. North Am.*, vol. 51, no. 6, pp. 1151–1167, 2018, doi: 10.1016/j.otc.2018.07.011.

[397] Duc-Phong NGUYEN, Tien-Tuan DAO, and Marie-Christine HO BA THO, “Valorization of REHAB_DEEPFACE,” Report, Université de Technologie de Compiègne.