Traitement dans le domaine chiffré par l'intelligence artificielle

Raghida El Saj

THÈSE DE DOCTORAT EN COTUTELLE INTERNATIONALE DE

L'UNIVERSITE DE BRETAGNE OCCIDENTALE

ECOLE DOCTORALE N°644
Mathématiques et Sciences et Technologies de l'Information et de la Communication en Bretagne Océane
Spécialité : Signal, Image, Vision, *Son*

ET
L'UNIVERSITE LIBANAISE

ECOLE DOCTORALE DES SCIENCES ET TECHNOLOGIE
Spécialité : *Ingénierie Informatique*

Par
Raghida EL SAJ

Traitement dans le domaine chiffré par l'intelligence artificielle

Thèse présentée et soutenue à Tripoli, le 19/12/2023
Unité de recherche : L@bISEN

Rapporteurs avant soutenance :

Rachid JENNANE	Professeur à l'Université d'Orléans
Maroun CHAMOUN	Professeur à l'Université Saint-Joseph de Beyrouth

Composition du Jury :

Président :	Mohamad DIAB	Enseignant chercheur à l'Université Rafic Hariri
Examinateurs :	Christian BROSSEAU	Professeur à l'Université Bretagne Occidentale
	Nesma SETTOUTI	Enseignante chercheuse à l'ISEN
Dir. de thèse (France) :	Ayman AL FALOU	Professeur à l'ISEN
Dir. de thèse (Liban) :	Mohamad KHALIL	Professeur à l'Université Libanaise

Invité(s) :

Ehsan SEDGH GOOYA	Enseignant chercheur à l'ISEN

Remerciements

À mes parents bien-aimés pour leur amour, leurs prières et leurs encouragements, en particulier ma mère, sans qui et sans ses sacrifices, je n'aurais pas atteint ce que je suis aujourd'hui.

À ma nouvelle petite famille, qui m'a comblée d'amour et m'a encouragée à continuer. À l'amour, aux conseils, au soutien et à la compréhension de mon mari, et à mon adorable petite Mona.

Je remercie tout particulièrement mes superviseurs, Ehsan SEDGH GOOYA, Ayman ALFALOU et Mohamad KHALIL, pour leur supervision et leurs suggestions constructives, qui ont contribué à la réussite de ce projet.

Je n'oublie pas de remercier tous les membres du jury qui ont accepté de considérer et d'examiner ce travail.

Par ailleurs, je tiens à remercier mes amis et collègues, ainsi que toutes les personnes qui, de près ou de loin, ont contribué à la réalisation de ce travail.

Table de Matières

INTRODUCTION GÉNÉRALE

L'intelligence artificielle est aujourd'hui bien plus qu'une simple innovation technologique ; elle représente une révolution qui façonne notre monde à de multiples niveaux. L'importance de l'intelligence artificielle dans la société moderne est indéniable, et ses implications touchent pratiquement tous les aspects de notre vie quotidienne. Même les informations privées telles que les données médicales, financières ou personnelles sont parfois utilisées par l'intelligence artificielle. Cela soulève des questions cruciales en matière de protection de la vie privée, de sécurité et d'éthique et implique la mise en place de politiques de confidentialité solides, de mesures de sécurité robustes, et d'une sensibilisation accrue aux questions éthiques liées à l'utilisation des données privées dans le domaine de l'intelligence artificielle. Ainsi, des réglementations sur la protection des données, telles que le règlement général sur la protection des données (RGPD) en Europe, sont apparues pour protéger l'utilisation des données personnelles privées. Il est également devenu impératif de développer des systèmes d'intelligence artificielle qui préservent la confidentialité des données privées.

L'objectif principal de cette thèse est de développer un système basé sur l'intelligence artificielle qui préserve la sécurité des données personnelles privées, en traitant des données cryptées. Le système développé assure la classification. Ce système peut être appliqué à une variété d'applications, mais les applications fondamentales de cette thèse sont la reconnaissance faciale et le contrôle d'accès.

La figure 1 montre le schéma général du système développé dans cette thèse. Le système peut être considéré comme composé de deux parties principales indépendantes, qui peuvent être utilisées séparément ou ensemble : la partie cryptage et la partie classification

Une nouvelle méthode de cryptage est développée dans cette thèse, basée sur la transformation en cosinus discrète (DCT). Cette méthode transforme les données claires dans le domaine fréquentiel, où elle utilise des clés de cryptage privées pour obtenir les informations cryptées finales. Le résultat final du cryptage est une image grise contenant peu d'informations, dont l'information d'origine est très difficile à déterminer. L'un des principaux avantages de cette méthode de cryptage est qu'elle ne nécessite aucune modification au niveau du réseau neuronal qui traitera les données cryptées. Le réseau de neurones

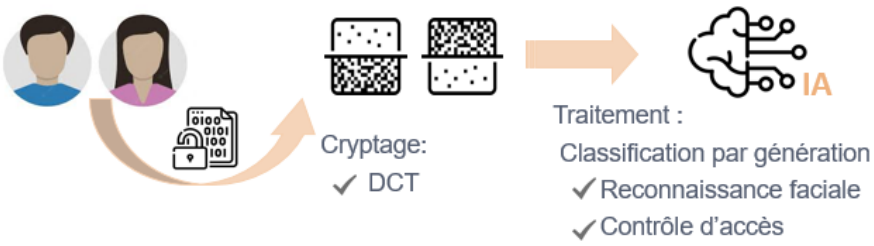


FIGURE 1 – Le schéma général du système développé.

peut être entraîné normalement et immédiatement sur les données cryptées, même sans augmentation du temps d'entraînement. La nouvelle méthode de cryptage peut être utilisée pour cacher n'importe quelle image, qu'elle soit ou non utilisée et traitée par le réseau de neurones.

Pour la partie classification de cette thèse, une nouvelle méthode de classification basée sur la génération est développée. Cette méthode de classification utilise n'importe quel réseau de neurones pouvant être entraîné à générer une sortie précise. Dans cette méthode, ce réseau de neurones est appelé "générateur". Le réseau générateur est entraîné à générer une sortie spécifique unique pour chaque classe. La décision concernant la classe est prise en fonction de la sortie générée. Cette méthode de classification nécessite la création d'une base de références contenant tous les résultats spécifiques souhaités pour chaque classe. Les générateurs testés dans cette thèse sont l'auto-encodeur et le pix2pix. La classification par génération est une méthode générale qui peut être utilisée pour n'importe quelle application, sur n'importe quel type de données, avec n'importe quel générateur.

La combinaison de ces deux parties crée un système d'intelligence artificielle qui traite les données cryptées, dans le but de préserver la confidentialité des données. Les données cryptées à l'aide de la nouvelle méthode de cryptage basée sur la DCT sont classées à l'aide de la méthode de classification par génération, afin de conserver leur confidentialité.

Ce rapport est organisé comme suit :

— **Chapitre 1 :** Ce chapitre présente l'état de l'art et explique le problème en détail. Des méthodes de préservation de la confidentialité des données sont présentées, ainsi que des méthodes de reconnaissance faciale.

— **Chapitre 2 :** Dans ce chapitre, la méthode de classification par génération est pré-

sentée en détail. Différentes applications sont testées, et les résultats sont discutés et interprétés.

— **Chapitre 3 :** La méthode de cryptage et le système global sont présentés dans ce chapitre. Des tests de robustesse sont effectués. En conclusion, les perspectives d'avenir sont exposées.

— **Conclusion générale et perspectives :** Ce manuscrit se termine par une conclusion générale qui résume l'ensemble des travaux réalisés et trace les perspectives d'avenir.

Chapitre 1

Besoin de sécurité

1.1 Introduction

L'Intelligence Artificielle (IA) est un processus d'imitation de l'intelligence humaine, basé sur la création et l'application d'algorithmes exécutés dans un environnement informatique dynamique. Son objectif est de permettre aux machines de penser et d'agir comme des êtres humains.

C'est dans les années 1950 que l'IA a été exposée, et la conférence de Dartmouth sur l'intelligence artificielle de 1956 [1] est considéré comme son point de départ officiel. Suite à cette conférence, pendant près de deux décennies, des succès significatifs ont été enregistrés dans le domaine de l'IA. Au cours de cette période, le célèbre programme informatique ELIZA [2] et le programme de résolution de problèmes généraux [3] ont été créés. En 1973, les dépenses élevées consacrées à la recherche en IA ont commencé à être fortement critiquées et le support financier accordé à ce type de recherche a diminué [4] ; cette période est considérée comme le premier hiver de l'IA.

Le développement du Réseau de Neurones (RN) convolutions multicouches hiérarchiques en 1980 [5] et l'invention de l'algorithme d'apprentissage par rétro-propagation en 1986 [6] ont eu un impact significatif et ont donné lieu à de hautes perspectives de réussite pour l'avenir. Toutefois, les succès réels ont été moins importants et les investissements dans le domaine de l'IA ont de nouveau diminué au début des années 1990. Cette période est considérée comme le deuxième hiver de l'IA.

Récemment, l'IA a fortement réapparu, notamment avec le développement et le grand succès de l'apprentissage automatique ou le Machine Learning (ML) et de l'apprentissage profond ou le Deep Learning (DL). Cette reprise est également liée à la disponibilité des données et à la réduction des coûts de calcul et de mémoire [7]. L'apprentissage profond est un type d'intelligence artificielle provenant de l'apprentissage automatique, où la machine est capable d'apprendre par elle-même à partir d'une grande quantité de données. Par conséquent, la préparation et la disponibilité des données sont essentielles au

développement d'un système intelligent performant. Les réseaux de neurones constituent la base des algorithmes d'apprentissage profond.

Dans le monde d'aujourd'hui, l'IA est devenue très importante. Elle est présente dans notre vie quotidienne, elle est appliquée dans un grand nombre de domaines, tels que la vision par ordinateur, la prédiction, l'analyse sémantique, le traitement du langage naturel et la recherche d'informations, afin de résoudre une grande variété de problèmes [8]. Il existe de nombreuses applications efficaces de l'apprentissage automatique dans la détection des objets [9] [10], les diagnostics médicaux [11] [12], la reconnaissance vocale [13] [14], manuscrite [15] et faciale [16] [17], la finance [18] et bien d'autres applications.

Certains domaines et applications utilisent des données privées et confidentielles, qui ne doivent pas être partagées ou diffusées partout. Il est donc impossible d'utiliser des solutions d'IA transportées dans des environnements en nuage[19], normalement non sécurisés. En outre, dans ces domaines, le problème de la disponibilité des données apparaît ; la collection des données devient très difficile et impose différentes normes et réglementations. D'où le besoin d'une solution qui préserve la confidentialité et la sécurité des données, tout en étant compatible avec les solutions basées sur l'IA. C'est ainsi qu'est apparu le développement des systèmes d'IA préservant la sécurité et la confidentialité des données.

Un autre domaine de préservation de la sécurité apparaît avec la forte domination de la digitalisation et sa large utilisation. Maintenir la sécurité des systèmes, établir des limites claires et contrôler l'accès des personnes afin d'autoriser ou de refuser l'accès à certaines fonctionnalités sont devenus des tâches fondamentales très importantes. D'où la nécessité de développer des systèmes de reconnaissance capables non seulement d'identifier et de classer les personnes, mais aussi d'identifier tout étranger qui n'a pas le droit d'accéder au système.

La première méthode traditionnelle et très simple de contrôle d'accès était basée sur des personnes physiques protégeant les bâtiments et autres biens physiques contre les entrées non autorisées. Puis, en 1961, les mots de passe numériques ont été utilisés comme méthode d'identification et d'authentification pour contrôler l'accès. Une personne possédant le bon mot de passe est une personne ayant accès. Plusieurs mots de passe peuvent être utilisés pour identifier différentes personnes ayant un accès.

Au fil du temps, les méthodes d'identification ont évolué pour être basées sur des informations connues ou des objets possédés. Les informations connues peuvent être des mots ou des phrases, des mots de passe graphiques, des numéros d'identification personnels

ou des réponses à des questions spécifiques. [20] [21] [22] [23] [24]. Les objets possédés peuvent être un jeton, une clé, un téléphone, une carte à puce ou une clé logicielle [25]. L'inconvénient de ces méthodes est qu'elles peuvent facilement être perdues, volées ou utilisées par des personnes non autorisées.

Aujourd'hui, les méthodes de contrôle d'accès biométriques sont de plus en plus utilisées en raison des avantages qu'elles présentent par rapport aux méthodes de contrôle traditionnelles. Les méthodes biométriques offrent un niveau de sécurité plus élevé du fait que les caractéristiques biométriques sont propres à chaque individu, d'où la possibilité de créer une correspondance unique entre une personne et un élément de données, ce qui rend extrêmement difficile aux utilisateurs non autorisés de les reproduire ou imiter [26] [27]. Ces méthodes sont plus pratiques et plus simples, les utilisateurs pouvant simplement présenter leurs caractéristiques biométriques, sans effort et sans avoir besoin de se rappeler des informations ou de présenter des jetons.

Les méthodes biométriques peuvent être basées sur la lecture d'empreintes digitales [28], la lecture de la rétine, la reconnaissance de l'iris [29], la reconnaissance vocale [30] ou la reconnaissance faciale [31]. La reconnaissance faciale est l'une des méthodes biométriques les plus utilisées.

La reconnaissance faciale ne nécessite aucune interaction humaine et elle est rapide et facile à utiliser [32]. Elle est généralement bien acceptée par les utilisateurs en raison de son caractère naturel et immédiat. Les utilisateurs sont habitués à la reconnaissance faciale dans leurs interactions quotidiennes et la trouvent plus facile à adopter que d'autres méthodes biométriques qui peuvent nécessiter un contact physique ou des actions spécifiques. De plus, le contrôle d'accès par la reconnaissance faciale ne nécessite pas d'équipement coûteux, il suffit d'une simple caméra pour capturer les visages afin de les traiter et de prendre des décisions.

Le traitement des images pour identifier les personnes et trouver les étrangers à partir de leurs visages est possible grâce à une variété de techniques, et ce domaine en particulier a été largement amélioré, notamment avec le développement de l'IA ; de nouvelles techniques et approches ont été développées et continuent de l'être.

Ce chapitre présente les deux axes de préservation de la sécurité. La première section présente les méthodes permettant de préserver la sécurité et la confidentialité des données, tandis que la seconde présente les méthodes permettant de préserver la sécurité en identifiant et en vérifiant les individus grâce à la reconnaissance faciale. Pour finir, la combinaison de ces deux axes, ainsi que l'objectif de la thèse, seront discutés dans la

conclusion.

1.2 Préservation de la confidentialité des données

Le fait que la disponibilité des données représente un élément très important dans le développement d'un système performant basé sur l'apprentissage profond et que, dans certaines applications, les données utilisées sont confidentielles et privées, le développement des systèmes préservant la confidentialité et la sécurité de ces données est devenue une priorité essentielle. L'une des méthodes permettant la protection des données et la préservation de leur confidentialité est le cryptage. Il était donc intéressant de développer des systèmes d'apprentissage automatique capables de traiter des données cryptées.

Combiner l'apprentissage automatique et la cryptographie est une idée qui existe déjà, les deux approches ont été rencontrées dans différentes applications. Des méthodes d'apprentissage profond ont été utilisées en stéganographie pour cacher et dissimuler des informations [33] [34] [35], ainsi que dans la cryptanalyse dans le but de développer de nouvelles méthodes d'attaque permettant la reconstitution des clés. Les méthodes d'apprentissage profond ont également été utilisées pour évaluer la sécurité d'un système cryptographie et déterminer la robustesse d'un algorithme de cryptage [36]. Ces méthodes ont également été appliquées pour créer des systèmes de cryptage ; le réseau de neurones a été utilisé comme une clé secrète de cryptage [37].

Les algorithmes d'apprentissage profond ont notamment été utilisés pour traiter et classer les données dans le domaine crypté afin de préserver la confidentialité et la sécurité des données. Plusieurs méthodes basées sur les réseaux de neurones ont été proposées et développées. Ces méthodes peuvent être classées en deux catégories : les méthodes de calcul et les méthodes de cryptage perceptuelles des images [38].

1.2.1 Méthodes de calcul

Les méthodes de calcul permettent aux RNs de traiter des données cryptées, sans avoir besoin de les décrypter et de connaître les informations en clair. Pour ce faire, des modifications spécifiques sont nécessaires et certaines limites sont imposées aux structures des RNs et à leurs fonctions d'activation. Il exige également que les données soient cryptées à l'aide d'une méthode de cryptage connues déjà existante comme le cryptage homomorphique (CH) ou le cryptage fonctionnel (CF). La différence entre CH et CF est

que le résultat du calcul des données cryptées en utilisant CH est crypté puisque CH évalue les données sans décryptage, alors que le résultat du calcul des données cryptées en utilisant CF est en clair puisque CF applique l'étape de décryptage pour évaluer le calcul des données cryptées [39].

Le CH évalue les données cryptées, de sorte que le résultat obtenu d'un calcul sur des données cryptées est le même que le cryptage du résultat obtenu du même calcul sur les données en clair [40]. CH ne supporte que l'addition et la multiplication comme opérations de calcul, donc uniquement les fonctions polynomiales. Soit *Enc* la fonction de chiffrement, * l'opération supportée par le CH, et x, y des exemples de données en clair, alors :

$$Enc(x) * Enc(y) = Enc(x * y) \tag{1.1}$$

De son côté, le CF d'une fonction f se compose de quatre algorithmes essentiels [41] :

— Setup : génère la clé publique mpk et la clé secrète principale msk de cryptage.
— KeyDrive : fournit la clé secrète sk_f de la fonction f à l'aide de msk.
— Cryptage : crypte le message x à l'aide de mpk.
— Décryptage : calcul f(x) en utilisant mpk, sk_f et le texte crypté de x généré lors de l'étape de *Cryptage*.

Dans ce qui suit, quelques méthodes de calcul, basées sur le cryptage homomorphique et sur le cryptage fonctionnel, sont présentées.

1.2.1.1 CryptoNets

En 2016, les CryptoNets [42] ont été proposés comme des RNs capables de classer des données cryptées par le CH. Comme le CH ne supporte que les fonctions polynomiales, toutes les fonctions du RN doivent l'être. Ainsi, certains ajustements nécessaires ont été imposés aux RNs pour créer les CryptoNets ; toutes les couches de pooling ont été remplacées par une couche de pooling à moyenne échelonnée dont la fonction est $\sum x$, et toutes les fonctions d'activation telles que *sigmoïde* et *ReLu* ont été remplacées par la fonction carrée x^2, à l'exception de la dernière fonction d'activation *sigmoïde* qui est nécessaire pour la phase d'apprentissage. Le RN ainsi modifié, totalement compatible avec le CH, est entraîné en utilisant des données non cryptées. Une fois l'entraînement terminé, la fonction d'activation sigmoïde est supprimée et les couches consécutives qui n'utilisent que des transformations linéaires sont réduites pour augmenter l'efficacité.

Pour utiliser les CryptoNets, l'utilisateur doit d'abord crypter ses données à l'aide

des clés de cryptage. Ensuite, il peut envoyer ses données cryptées au réseau CryptoNets pour être classées et ainsi recevoir un résultat crypté. Enfin, l'utilisateur doit décrypter le résultat obtenu à l'aide de la clé de cryptage secrète pour atteindre le résultat final non crypté.

Bien que les CryptoNets réalisent de bonnes performances, ils présentent de nombreuses difficultés et limitations. Le traitement des données avec les CryptoNets prend beaucoup de temps, et une prédiction prend 250 secondes. De plus, le réseau doit d'abord être entraîné sur des données non cryptées, puis testé sur des données cryptées, ce qui ne résout pas complètement le problème de la sécurité des données. En outre, les CryptoNets utilisent un réseau de neurones à double profondeur, qui n'est pas considéré comme un réseau profond, d'où la limitation de la profondeur du réseau à utiliser.

1.2.1.2 CryptoDL

En 2017, une solution permettant aux réseaux de neurones profonds de traiter des données chiffrées a été proposée [43]. Cette technique se compose de deux éléments de base : les réseaux de neurones convolutifs (RNC) et le CH, précisément le CH à niveaux.

Pour rendre le RNC compatible avec le CH, les couches de pooling ont été remplacées par des couches de pooling à moyenne échelonnée, et une nouvelle méthode a été conçue pour approximer les fonctions d'activation les plus courantes (*ReLu*, *sigmoïde*, et *tangh*) avec des polynômes de faible degré. Le polynôme à degré élevé permet d'obtenir de meilleures performances, mais à un coût de calcul élevé ; par conséquent, seuls les polynômes de degré deux et trois ont été utilisés.

La technique d'approximation proposée est basée sur la dérivée de la fonction d'activation au lieu de la fonction d'activation elle-même, en utilisant des polynômes de degré trois. Cette méthode a été comparée avec l'analyse numérique, les séries de Taylor, les polynômes de Tchebychev standard et les polynômes de Tchebychev modifiés pour l'approximation de la fonction *ReLu*, et a permis d'obtenir la meilleure approximation. En outre, il a été prouvé que le comportement de l'approximation polynomiale de la fonction *ReLu* est robuste aux modifications de la structure du RNC. De plus, les différentes fonctions d'activation, *ReLu*, *sigmoïde*, et *tangh*, avec leur remplacement polynomial ont été comparées et la meilleure précision a été obtenue en utilisant la fonction d'activation *ReLu*.

CryptoDL doit être entraîné sur des données non cryptées. Après l'apprentissage, les nouvelles données à traiter doivent être cryptées. Les résultats de CryptoDL sont affectés

par la base utilisée et par l'application. Tout comme pour les CryptoNets, CryptoDL ne résout pas totalement le problème de la préservation de la sécurité des données, tant que le processus d'apprentissage doit être appliqué à des données non cryptées.

1.2.1.3 CryptoNN

En 2019, CryptoNN [44] a été proposé comme un modèle qui supporte à la fois les phases d'entraînement et de test sur des données cryptées. Ce modèle repose sur un calcul matriciel sécurisé basé sur le cryptage fonctionnel.

Un schéma de cryptage fonctionnel CFPI pour une fonction de produit interne $f(x, y)$ a été adopté, où n est la longueur des vecteurs de données x et y. Un deuxième schéma de cryptage fonctionnel CFOB pour les opérations de base $f_\Delta(x, y)$ a été proposé, où Δ peut être l'addition, la soustraction, la multiplication, ou la division.

$$f(x, y) = \sum_{i=1}^{n} (x_i y_i) \tag{1.2}$$

$$f_\Delta = x \Delta y \tag{1.3}$$

Le modèle CryptoNN se compose de trois unités, comme le montre la figure 1.1 :

— L'autorité : génère la clé publique mpk, la clé secrète principale msk, et la clé secrète de la fonction sk_f (étape KeyDrive). CFOB a différentes approches pour la génération de sk_f ; la clé secrète de la fonction sk_{f_Δ} est calculée selon Δ.

— Le client : prépare et crypte les données - les entrées (x) et les étiquettes (y) - à l'aide de mpk et les envoie au serveur (étape Cryptage). Les étiquettes doivent d'abord être codées à l'aide de la méthode encodage à chaud, puis mises en correspondance avec un vecteur aléatoire r dont les composantes sont r_i.

— Le serveur : entraîne et teste le RN en utilisant les données reçues des clients. Ayant les données d'entrée et la première couche cachée dans le processus de propagation avant, et les étiquettes et la couche de sortie dans le processus de rétro-propagation, le serveur obtient de l'autorité le sk_f correspondant à la fonction spécifique et décrypte ensuite le résultat de la fonction (étape de Décryptage). Le serveur peut poursuivre normalement les processus de rétro-propagation et de propagation avant. La sortie du réseau est p_i, qui est la probabilité que les données d'entrée x appartiennent à la classe i.

Le système proposé comporte donc deux cycles de calcul sécurisé : au début du proces-

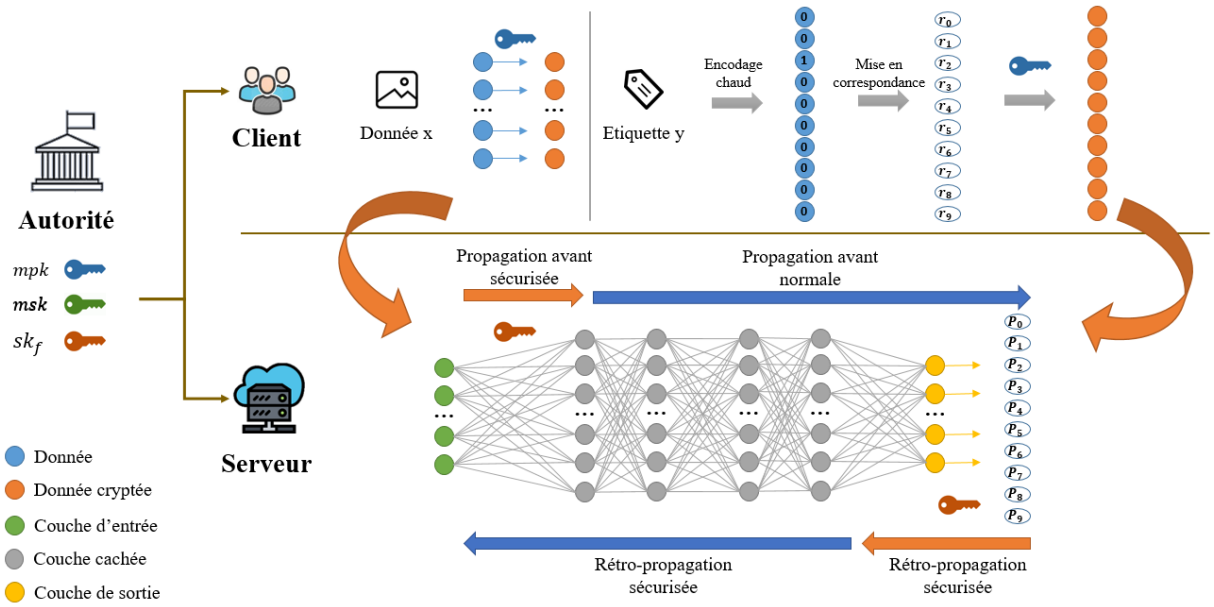


FIGURE 1.1 – CryptoNN.

sus de propagation avant, appelé propagation avant sécurisée, et au début du processus de rétro-propagation, appelé rétro-propagation / évaluation sécurisée. Un cas pratique de CryptoNN, CryptoCNN, utilisant l'architecture LetNet-5 qui comprend cinq couches cachées, a été créé et utilisé. La propagation avant sécurisée a lieu dans la première couche cachée, soit la couche convolutive, tandis que la rétro-propagation sécurisée a lieu dans la couche de sortie.

Le CryptoCNN peut être considéré comme une solution au problème de la sécurité des données, car il permet d'entraîner et de tester le RN sur des données cryptées, mais il pose également un problème de temps pour entraîner ce réseau. Son entraînement nécessite un long temps avec l'intégrité du calcul matriciel sécurisé.

1.2.2 Méthodes perceptuelles

Les méthodes perceptuelles protègent l'information visuelle en générant des images incompréhensibles, directement appliquées aux algorithmes de traitement d'images et aux RNs sans aucune modification. De nombreuses méthodes perceptuelles ont été proposées, dont certaines peuvent être appliquées aux algorithmes d'apprentissage automatique traditionnels tels que les machines à vecteurs de support et les forêt d'arbres décisionnels [45], [46]. De même, des méthodes de cryptage d'images perceptuelles ont été établies

pour être utilisées sur des RNs sans imposition de contraintes ou de modifications sur ces réseaux ni sur leurs fonctions d'activation, contrairement aux méthodes de calcul.

Dans ce qui suit, quelques méthodes de cryptage perceptuelles des images sont présentées.

1.2.2.1 Le schéma de Tanaka

Le schéma de Tanaka [47], proposé en 2018, est basé sur le cryptage par blocs. L'image claire, composée de trois canaux RVB de 8 bits, doit d'abord être divisée en blocs de dimensions $M \times M \times 3$ chacun, comme le montre la figure 1.2. Ensuite, chaque bloc doit être divisé en valeurs de 4 bits des pixels supérieurs et inférieurs pour former des blocs d'image à 6 canaux. Pour crypter les informations portées par les pixels, leurs valeurs doivent être inversées et réarrangées de manière aléatoire à l'aide de la clé secrète d'inversion K_{inv} et de la clé secrète de ré-arrangement K_{rear}. On obtient enfin une partie de l'image cryptée en reformant les blocs à 6 canaux en blocs à 3 canaux. L'image cryptée est le regroupement de toutes les parties cryptées. Pour que les images cryptées puissent être traitées par le RN et afin de réduire l'influence du cryptage des images, un réseau d'adaptation doit être ajouté avant le RN à utiliser.

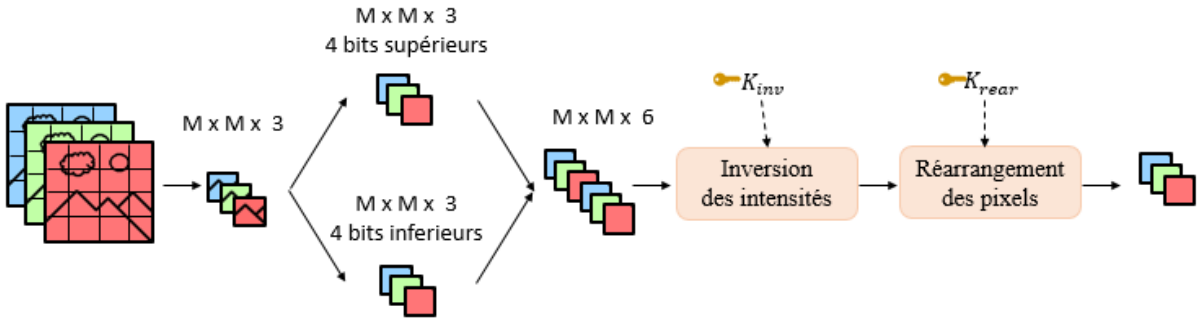


FIGURE 1.2 – Le schéma de Tanaka.

En 2020, une méthode de brouillage d'image par blocs a été proposée pour augmenter le niveau de sécurité des images protégées visuellement [48]. Cette méthode est connue sous le nom de cryptage apprenable étendu. Dans cette méthode, après avoir divisé l'image en $M \times M$ blocs, les positions des blocs sont réarrangées (Voir la figure 1.3), ensuite les pixels dans chaque bloc le sont aussi. Enfin, les blocs sont concaténés et l'image cryptée

est obtenue. L'augmentation du niveau de sécurité de cette méthode cause une diminution de la précision de la classification.

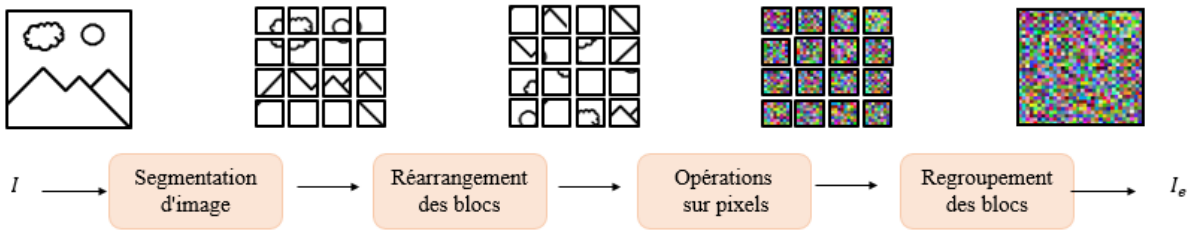


FIGURE 1.3 – Le cryptage apprenable étendu.

1.2.2.2 Le cryptage d'images basé sur les pixels

Une méthode de cryptage d'images basée sur les pixels qui prend en compte l'augmentation des données dans le domaine crypté a été proposée en 2019 [49] [50]. Avec cette méthode, l'augmentation des données peut être effectuée par l'utilisateur avant le cryptage ou par le serveur après le cryptage.

Pour générer une image cryptée I_e à partir d'une image couleur a trois canaux (rouge, vert et bleu) $I = \{I_R, I_V, I_B\}$ avec n pixels, l'image I doit d'abord être divisée en pixels (Voir la figure 1.4). Ensuite, en fonction d'un entier binaire aléatoire $r(i)$, une transformation Négatif-Positif (NP) doit être appliquée individuellement à chaque pixel p des trois canaux de couleur pour obtenir un pixel transformé p'. L'entier $r(i)$ est généré en utilisant un ensemble de clés secrètes $K_{NP} = \{K_R, K_G, K_B\}$, utilisées respectivement pour I_R, I_G et I_B, où i est le $i$ème pixel de I. Si r(i) est égal à zéro, la valeur du pixel est conservée telle quelle, alors $p' = p$. Sinon, la transformation NP doit être appliquée au pixel p formé par L bits :

$$\begin{aligned} p' &= p && si \quad r(i) = 0 \\ p' &= p \oplus (2^L - 1) && si \quad r(i) = 1 \end{aligned} \tag{1.4}$$

Enfin, les trois composantes de couleur de chaque pixel peuvent être réarrangées en utilisant la clé secrète K_r. Ainsi, la clé de cryptage secrète devient $K = \{K_{NP}, K_r\}$.

Deux possibilités de clés de cryptage existent pour générer des images d'apprentissage et de test cryptées :

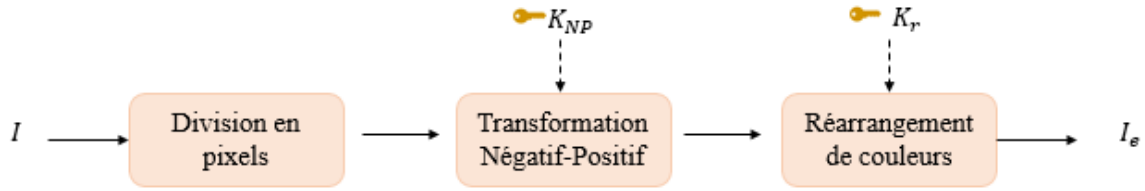


FIGURE 1.4 – Le Cryptage d'images basé sur les pixels.

— Même clé de cryptage : toutes les images d'entraînement et de test sont cryptées à l'aide de la même clé de cryptage K.
— Clés de cryptage différentes : des clés différentes sont attribuées indépendamment aux images d'apprentissage et aux images de test.

Comme pour la méthode de Tanaka, un réseau d'adaptation composé de 1×1 couches de convolution a été proposé pour rendre les images cryptées compatibles avec le RN.

1.2.2.3 Le schéma de transformation d'images basé sur le GAN

En 2020, un réseau de transformation d'images (TN-GAN) a été développé à partir du réseau antagoniste génératif - Generative Adversarial Network (GAN), afin de préserver la confidentialité des données lors de l'utilisation des RNs profonds [51]. Ce réseau de transformation $h_p(.)$ a été utilisé pour protéger les images d'entraînement et les images de test en les transportant du domaine clair A vers un domaine B visuellement protégé. Pour obtenir $h_p(.)$, une translation d'image à image non appariée basée sur un réseau antagoniste génératif à cycle-consistant (cycle-GAN) a été entraînée et utilisée.

Le cycle-GAN se compose de deux GANs : deux réseaux génératifs G_{AB} et G_{BA} et deux réseaux discriminants D_A et D_B (voir la figure 1.5). Son objectif est de transporter des images d'un domaine à l'autre à l'aide des générateurs et de s'assurer que l'image générée dans le nouveau domaine y correspond parfaitement à l'aide des discriminants. Dans cette application, les deux domaines sont le domaine clair A et le domaine visuellement protégé B. En effet, G_{AB} transporte des images x du domaine A vers le domaine B, et D_B distingue la différence entre les images nouvellement générées $G_{AB}(x)$ et les images réelles du domaine B x_p. De même, G_{BA} transporte des images x_p du domaine B vers le domaine A, et D_A distingue la différence entre les images nouvellement générées $G_{BA}(x_p)$ et les images réelles du domaine B x.

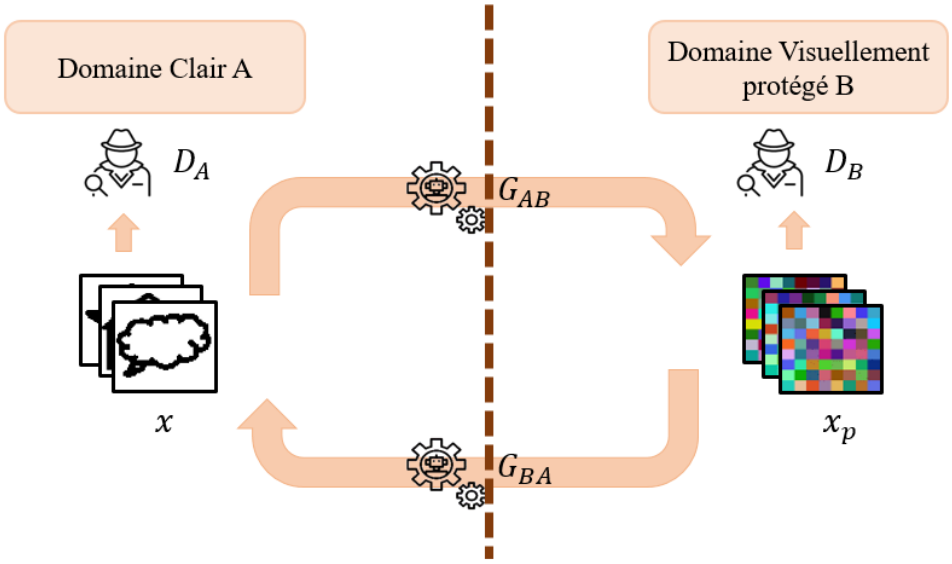


FIGURE 1.5 – Cycle-GAN.

Le réseau génératif G_{AB} du cycle-GAN est utilisé comme le réseau de transformation $h_p(.)$. Ce réseau est entraîné en utilisant des images X avec leurs étiquettes correspondantes Y et un ensemble d'images protégées préliminaires P, comme illustré à la figure 1.6. Les images P sont générés à partir d'un modèle pré-entraîné $h_\theta(.)$ en utilisant X et Y dans un processus de protection visuelle préliminaire. La sortie $X_p = h_p(X)$ de $h_p(.)$ est une image protégé visuellement, qui peut être utilisée directement pour entraîner un autre RN afin de préserver la confidentialité des données. Dans l'application, le VGG-13 a été utilisé comme le réseau pré-entraîné $h_\theta(.)$, le U-Net a été utilisé comme le réseau de transformation $h_p(.)$, et le ResNet-18 a été utilisé comme le RN de classification.

1.2.2.4 Le schéma de transformation d'images basé sur un modèle

Un réseau de transformation d'images entraîné par un modèle (TN-modèle) a été proposé encore en 2020 [52] [53]. Le réseau de transformation $h_\theta(.)$, qui peut être accessible à tout le monde, est entraîné en parallèle avec un modèle de classification ψ, pour générer des images visuellement protégées (Voire figure 1.7).

Le modèle de classification ψ, généralement installé dans le nuage, est entraîné à partir des images X en clair et de leurs étiquettes correspondantes Y. Un sous-ensemble des images d'entraînement ($X' \subseteq X$) dont l'ensemble des étiquettes correspondantes est Y', est envoyé à l'utilisateur pour entraîner le réseau de transformation $h_\theta(.)$. La sortie du

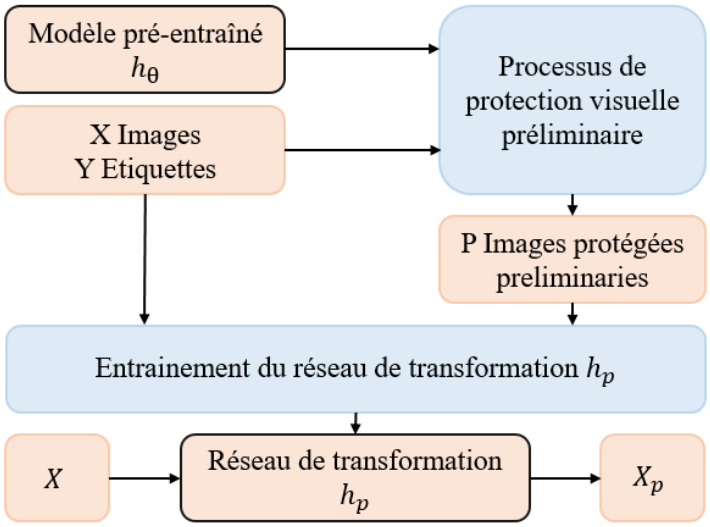


FIGURE 1.6 – Le processus d'apprentissage de $h_p(.)$ dans la méthode TN-GAN.

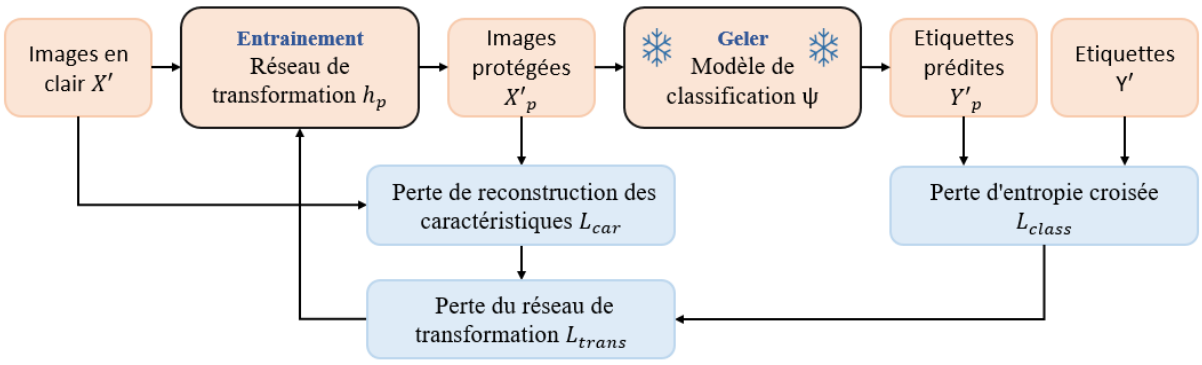


FIGURE 1.7 – Le processus d'apprentissage de $h_p(.)$ dans la méthode TN-modèle.

réseau de transformation, $X'_p = h_\theta(X)$, est en ensemble des images visuellement protégées qui doivent être renvoyées au classificateur afin qu'il tente de les classer correctement. Les résultats correspondants prédits par le modèle de classification sont Y'_p.

La fonction de perte du réseau de transformation L_{trans} doit être minimisée.

$$L_{trans}(x_i, x_{pi}, y_i) = L_{class}(x_{pi}, y_i) - \alpha \ . \ L_{car}(x_i, x_{pi}), \tag{1.5}$$

où i est un nombre entier compris entre 1 et m, m est le nombre d'images dans X'. La perte de classification L_{class} est une fonction de perte d'entropie croisée calculée en utilisant $\hat{y_i} = \psi(\hat{x_i})$ et y_i. α est un coefficient de L_{car} qui est la perte de reconstruction des caractéristiques calculée à l'aide du 'feature map' de l'image originale $\phi_k(x_i)$ et du 'feature map' de l'image reconstruite protégée visuellement $\phi_k(\hat{x_i})$. C_k, H_k, W_k représentent les dimensions du 'feature map' $\phi_k(.)$.

$$L_{car}(x_i, x_{pi}) = \frac{1}{C_k \times H_k \times W_k} ||\phi_k(x_{pi}) - \phi_k(x_i)||_2^2. \tag{1.6}$$

1.2.3 Discussion

Avec la large diffusion et l'efficacité des solutions basées sur les réseaux de neurones, et avec le souci important de la préservation de la confidentialité des données, il était nécessaire de développer des réseaux de neurones préservant la confidentialité. Les méthodes préservant la confidentialité des données permettent l'utilisation et le traitement des données privées par les réseaux de neurones sans être obligé de diffuser les informations. Les méthodes préservant la confidentialité des données se concentrent sur la classification des données cryptées, et atteignent des performances intéressantes. Ces méthodes se divisent en deux catégories : les méthodes de calcul et les méthodes perceptuelles.

Les méthodes de calcul visent à utiliser les RN sur des données cryptées par des méthodes de cryptage déjà existantes et connues, telles que le cryptage homomorphique et le cryptage fonctionnel. Ces méthodes permettent d'obtenir de bons résultats, mais présentent certaines limites en termes de profondeur du RN utilisé, de coût de calcul élevé, de temps de prédiction et d'entraînement, et d'interdiction de l'entraînement sur des données cryptées.

CryptoNets et CryptoDL modifient le RN pour le rendre compatible avec la méthode de cryptage CH utilisée. Les principales différences entre CryptoNets et CryptoDL sont les fonctions d'activation choisies et la technique d'approximation utilisée. Le choix de la

fonction d'activation et la technique d'approximation appliquée ont un impact significatif sur la performance de la classification du RN et sur le temps de prédiction. La latence de prédiction est causée par la complexité de calcul ; toutes les fonctions doivent être calculées à l'aide d'additions et de multiplications imbriquées. Elle est également liée à la grande dimension des données cryptées par rapport aux données non cryptées ; les données cryptées sont une à trois fois plus grandes que les données non cryptées. Cela implique le transfert de grandes quantités de données.

CryptoNN permet au RN de traiter des données tout en préservant leur confidentialité et leur sécurité grâce au calcul matriciel sécurisé basé sur le CF, qui ne nécessite pas la modification des fonctions et de la structure du RN. Tandis que CryptoNets et CryptoDL permettent la classification de données cryptées à l'aide d'un RN précédemment entraînés sur des données non cryptées, le CryptoNN permet à la fois l'entraînement et le test sur des données cryptées. Par contre, la difficulté de CryptoNN réside dans la nécessité de communiquer fréquemment avec l'autorité afin de générer et d'obtenir les clés correspondantes. Cette difficulté et les calculs cryptographies complexes qu'elle implique font que CryptoNN nécessite un temps d'apprentissage beaucoup plus long que le réseau original.

Toutes ces méthodes de calcul n'ont été testées que sur des bases de données simples, telles que MNIST et CIFAR-10, de sorte que leur variabilité n'est pas encore certaine. Et malgré la simplicité des bases de données et l'utilisation des réseaux peu profonds, le calcul était très complexe et son coût était très élevé. Cela souligne la difficulté d'appliquer ces méthodes aux réseaux de neurones profonds convolutifs et de résoudre des problèmes plus réalistes.

D'autre part, les méthodes de cryptage perceptuelles transforment ou cryptent une image d'une manière qui la rend incompréhensible pour les humains, mais toujours apprise par les RNs. Les images qui en résultent sont des images visuellement protégées qui ne contiennent aucune information claire. Contrairement aux méthodes de calcul, ces méthodes peuvent être appliquées à n'importe quel type de réseau, sans limitation d'architecture ou de fonctions d'activation. Cependant, un nouveau problème apparaît, lié à la clé de cryptage utilisée pour cacher l'information, et à la possibilité de régénérer les images initiales et de récupérer les informations privées.

Le schéma de Tanaka et la méthode de cryptage d'images basée sur les pixels permettent à la fois d'entraîner et de tester le RN sur des images cryptées. Ils utilisent des réseaux d'adaptation pour réduire l'influence du cryptage d'images, tandis que l'analyse de ces réseaux peut permettre de comprendre le processus de cryptage, de sorte que les

images visuellement protégées peuvent être facilement reconstruites par des adversaires. L'utilisation d'un réseau d'adaptation n'est donc pas une solution parfaite, malgré les avantages offerts.

Les deux méthodes, le schéma de Tanaka et la méthode de cryptage basée sur les pixels, ont un faible coût de calcul et appliquent des fonctions de transformation simples aux images initiales. Par conséquent, les adversaires peuvent déduire le processus de cryptage à l'aide de méthodes d'attaque par texte chiffré uniquement, surtout lorsque la même clé de cryptage est utilisée pour crypter toutes les images. Toutefois, la méthode de cryptage d'images basée sur les pixels présente un espace de clés plus grand que le schéma de Tanaka si l'image est plus grande que 11×11 pixels, et elle autorise l'utilisation des différentes clés de cryptage pour chaque image tout en maintenant la performance de la classification. Cette méthode permet également l'augmentation des données dans le domaine crypté. Les performances obtenues en appliquant l'augmentation des données avant ou après le cryptage sont très proches.

Contrairement au système de Tanaka et à la méthode de cryptage basée sur les pixels, TN-GAN et TN-modèle n'utilisent pas de simples fonctions de transformation pour générer des images visuellement protégées, mais utilisent des réseaux de transformation. Ces réseaux doivent être entraînés, ce qui augmente le coût de calcul. Le coût de calcul du TN-GAN est plus élevé que celui du TN-modèle parce que le TN-GAN doit entraîner un cycle-GAN composé de deux réseaux génératifs et de deux réseaux discriminants pour obtenir son réseau de transformation, alors que le TN-modèle n'entraîne qu'un seul réseau. Cependant, le TN-modèle ne permet pas d'entraîner le RN sur des données cryptées, alors que le TN-GAN le permet.

1.3 Reconnaissance faciale

La reconnaissance faciale est l'un des domaines de recherche les plus actifs dans le domaine de la vision par ordinateur et de la reconnaissance des formes. Bien que les systèmes existants d'apprentissage automatique et de reconnaissance des visages présentent un certain degré de maturité et de bonnes performances, ils sont encore loin des capacités du système visuel humain, à cause des influences des diverses conditions imposées dans les applications réelles, telles que l'éclairage, la posture, les expressions faciales et bien d'autres [31] [54].

Un système de reconnaissance faciale peut fonctionner en mode de vérification ou en

mode d'identification [55]. La vérification est un processus un à un qui consiste à confirmer ou à infirmer l'identité déclarée d'une personne. Par contre, l'identification est un processus un à plusieurs qui consiste à déterminer l'identité d'une personne généralement en comparant ses données avec toutes les données identifiant les personnes dans la base de données.

La reconnaissance faciale commence par la détection des visages dans l'image, afin de déterminer si l'image contient ou non des visages et de les localiser si nécessaire. Ensuite, les caractéristiques des visages doivent être extraites pour créer le vecteur de signature représentant chaque visage. Enfin, la reconnaissance faciale, la décision, peut être effectuée à partir des vecteurs de signature [56]. Selon les étapes d'extraction des caractéristiques et de décision utilisées, les méthodes de reconnaissance faciale peuvent être divisées en quatre approches différentes [54] : approche holistique, approche locale, approche par texture locale, et approche par apprentissage profond.

1.3.1 Approche holistique

L'approche holistique ou l'approche du sous-espace considère l'ensemble du visage comme une entité unique pour l'identification. Elle prend en compte la structure et l'aspect global du visage et représente l'image du visage par un vecteur obtenu en alignant les différentes lignes de l'image. Le vecteur est implémenté dans un espace de faible dimension, afin de réduire sa dimension. L'étape de décision est effectuée en projetant le visage de l'image test et en calculant la mesure de la distance avec toutes les classes dans le sous-espace. Selon la représentation du sous-espace, les méthodes holistiques sont divisées en méthodes linéaires et non linéaires.

Eigenface est l'une des approches holistiques linéaires les plus connues [57] [58]. Elle est basée sur la technique de l'analyse en composantes principales (ACP). L'objectif de l'ACP est de réduire la taille de l'espace de données. L'ACP à noyau est une amélioration de la méthode eigenface qui utilise la technique de la méthode à noyau. C'est une méthode holistique non linéaire [59]. Fisherface est une autre méthode holistique linéaire qui utilise l'analyse discriminante linéaire (ADL) au lieu de l'ACP pour réduire l'espace d'image à haute dimension [60], [61]. La différence réside dans le fait que l'ACP est une technique supervisée, tandis que l'ADL est une technique non supervisée. La transformée en cosinus discrète (TCD) et la transformée en ondelettes discrète (TOD) sont d'autres techniques qui ont été employées dans la compression d'images et la sélection de caractéristiques pour l'analyse faciale [62] [63].

Les méthodes holistiques offrent une reconnaissance robuste dans différentes conditions d'éclairage et d'expressions faciales. Bien que ces méthodes permettent une meilleure réduction de dimension et améliorent le taux de reconnaissance, elles ne sont pas invariantes aux translations et aux rotations [31].

1.3.2 Approche local

Les points de repère du visage sont utilisés pour enregistrer les caractéristiques faciales, la normalisation des expressions et la reconnaissance de positions définies. Les points de repère les plus couramment utilisés sur le visage sont le bout du nez, le bout des yeux, le bout des coins de la bouche, les sourcils, le milieu de l'iris, le haut de l'oreille, les narines et le nez. La distribution des points de repère est utilisée dans l'approche locale dans la structure des règles heuristiques impliquant des distances, des angles et des régions.

La correspondance graphique élastique (CGE) est une méthode connue de reconnaissance des visages basée sur l'utilisation de points de repère. Cette méthode est une application de la construction dynamique d'arcs pour la détection d'objets, qui fournit une approche réaliste de la reconnaissance des visages [64] [65]. L'appariement de grappes élastiques est une extension de la CGE [66].

L'un des principaux inconvénients de toutes les méthodes basées sur la géométrie est la nécessité de disposer d'images faciales parfaitement alignées. Pour obtenir des résultats précis, les images du visage doivent être alignées afin de s'assurer que tous les points de référence correspondent aux caractéristiques appropriées du vecteur de caractéristiques. En général, l'alignement des images faciales est effectué manuellement, ce qui implique souvent un changement d'échelle anisotrope. L'obtention d'un alignement automatique optimal est considérée comme une tâche difficile dans ce contexte [54].

1.3.3 Approche par texture locale

L'approche de la texture locale est basée sur les importantes informations que porte la texture. Il s'agit d'une technique utilisée pour extraire des caractéristiques de n'importe quel objet et qui joue un rôle important dans la reconnaissance des formes et la vision par ordinateur. Le motif binaire local (LBP) est la technique la plus couramment utilisée, et elle a été largement appliquée à la reconnaissance des visages. La quantification de phase locale, un opérateur nouvellement introduit, a également été utilisée dans la reconnaissance des visages pour résoudre le problème de la reconnaissance des visages

flous.

Ces méthodes se caractérisent par la complexité de la détection automatique des caractéristiques pertinentes et l'incapacité à faire de la discrimination. De plus, elles rencontrent des difficultés dans les situations suivantes : variations de la posture, faible résolution des images, variations d'expression faciale et conditions d'éclairage variables [54].

1.3.4 Approche par apprentissage profond

Avec l'utilisation et le développement étendus de l'IA, différentes méthodes ont été utilisées pour la reconnaissance des visages. Certaines méthodes sont basées sur l'une des approches précédentes et utilisent l'IA, comme les machines à vecteurs de support [67] [68], les k-plus proches voisins [69] [70], les k-moyennes [71] [72] et l'apprentissage profond, pour la classification et la prise de décision.

Les approches d'apprentissage profond sont utilisées non seulement comme classifieur classificateur multiclasses pour séparer les différentes identité faciales. Les réseaux de neurones profonds, en particulier les réseaux de neurones convolutionnels, sont entraînés pour apprendre des caractéristiques faciales approfondies plus discriminantes. Certaines méthodes se concentrent sur l'extraction de caractéristiques à partir de différentes régions du visage, tandis que d'autres se focalisent sur l'extraction de caractéristiques à partir d'images faciales avec des variations d'apparence non frontales [73] [74] [75]. De nombreux travaux intègrent des concepts de l'apprentissage métrique, en combinant différentes fonctions de perte ou en utilisant des méthodes de fonction de perte efficaces [76] [77] [78]. Parallèlement, d'autres approches mettent en œuvre des fonctions d'activation appropriées pour améliorer les performances [79] [80].

1.4 Conclusion

Le monde d'aujourd'hui pose de nouveaux défis en matière de sécurité. Avec la domination des techniques intelligentes et de la numérisation, le maintien de la sécurité des systèmes devient une tâche importante. Le développement de solutions d'intelligence artificielle a permis de réaliser des systèmes de contrôle d'accès basés sur des données que l'individu ne peut ni oublier ni perdre, des données biométriques qui lui accompagnent tout au long de sa vie, qui lui sont propres et uniques et qu'il est donc difficile d'imiter.

La reconnaissance faciale est l'une des méthodes de contrôle d'accès biométrique, lar-

gement développée par l'intelligence artificielle et les techniques d'apprentissage profond. La reconnaissance faciale utilise les images des visages des personnes, qui sont des données privées pouvant être utilisées et exploitées pour des applications sans leur autorisation. D'où le problème de la disponibilité des données. D'autre part, les systèmes d'intelligence artificielle et les réseaux neuronaux nécessitent de très grandes quantités de données, afin d'obtenir de bonnes performances.

Le problème de la disponibilité des données privées pour l'entraînement des systèmes d'intelligence artificielle peut être résolu en développant des systèmes préservant la confidentialité des données. Grâce aux méthodes de cryptage perceptuelles des images, le réseau peut être entraîné et utilisé sur des données cryptées sans qu'il soit nécessaire de connaître les informations en clair.

l'objectif de cette thèse est de développer un système pour le contrôle d'accès basant sur la reconnaissance faciale tout en préservant la sécurité des données de bout en bout. Une nouvelle approche d'identification basée sur la génération de données est appliquée. La confidentialité des données est préservée par l'application d'une nouvelle méthode de cryptage basée sur la transformée en cosinus discrète (DCT). Toutes les images sont cryptées, puis envoyées au système pour l'entraîner et l'utiliser ultérieurement. Le système, constitué d'un réseau capable de générer des données, génère une image code que seul l'utilisateur peut comprendre, en la comparant à une base de référence déjà créée. Différents réseaux ont été tester.

Chapitre 2

Classification par génération

2.1 Introduction

Le domaine de l'intelligence artificielle est en évolution et développement continus. De nouvelles approches, techniques et applications apparaissent et se développent en permanence pour résoudre des problèmes complexes dans divers domaines. Les applications courantes de l'IA comprennent la classification, la régression, la génération, le traitement du langage naturel, l'optimisation, et bien d'autres encore.

Les algorithmes d'IA peuvent être utilisés dans la classification des données, où le modèle tente de prédire la classe correcte pour un élément de données, après avoir été entraîné sur une base de données avec ses classes correspondantes [81]. Dans la littérature, différents algorithmes ont été développés pour classer les données, tels que les machines à vecteurs de support, les k-plus proches voisins et les réseaux de neurones [82] [83]. Les réseaux de neurones, en particulier les réseaux de neurones profonds, ont attiré une attention particulière grâce à leurs performances exceptionnelles dans les tâches de classification.

La généralisation, l'apprentissage et le test sont bien meilleurs dans les réseaux neuronaux que dans les autres méthodes. Les réseaux neuronaux peuvent capturer des modèles et des caractéristiques complexes dans les données, ce qui les rend adaptés à des tâches complexes telles que la reconnaissance d'images, la compréhension du langage naturel et le traitement de la parole. Ils peuvent effectuer un apprentissage de bout en bout, c'est-à-dire qu'ils peuvent prendre des données brutes en entrée et produire des résultats significatifs en sortie, sans qu'il soit nécessaire de procéder à une ingénierie approfondie des caractéristiques. Ce qui simplifie le processus de développement et peut conduire à des modèles plus précis. De plus, les réseaux de neurones gèrent mieux les grandes bases de données. Les réseaux de neurones convolutifs sont largement utilisés pour la classification d'images [84].

Les algorithmes d'IA sont également capables de générer de nouvelles données en apprenant à partir de données existantes. Grâce aux connaissances acquises, le modèle

génératif crée de nouvelles données uniques. Les réseaux génératifs antagonistes (GAN) [85] et les auto-encodeurs variationnels (VAE) sont les modèles génératifs les plus communs [86].

Une nouvelle méthode de classification basée sur la génération de données par réseaux de neurones est développée et présentée dans ce chapitre. Cette méthode est basée sur des modèles capables d'être entraînés pour générer des données. Cette génération de données est déterminée par une base de références créée précédemment. La classification est finalement réalisée en comparant les données nouvellement générées avec les données de la base de références. Deux réseaux, pix2pix et auto-encodeur (AE), ont été utilisés pour la génération de données et ont été testés.

Ce chapitre présente d'abord réseaux de neurones puis les GAN, en particulier pix2pix, et les auto-encodeurs. Ensuite, la nouvelle approche de classification incluant la création de la base de références, les réseaux utilisés et la comparaison pour prendre la décision sont présentés. Enfin, les résultats obtenus sur différentes applications ainsi que leur interprétation concluent ce chapitre.

2.2 Réseaux de neurones

Les réseaux de neurones sont l'un des algorithmes d'intelligence artificielle, en particulier un type d'apprentissage profond, qui sont inspirés par la structure et le fonctionnement du cerveau humain, imitant la façon dont les neurones biologiques se signalent les uns aux autres.

Un réseau de neurones est constitué de nœuds interconnectés, organisés en couches, qui traitent et transforment les données. Chaque nœud ou neurone reçoit des données X_i, les traite et produit une sortie en appliquant une fonction mathématique appelée la fonction d'activation (AF) (Voir la figure 2.1). Des poids (W_i)sont appliqués aux données lorsqu'elles traversent les synapses pour atteindre le neurone. Ils déterminent la force des connexions et influencent l'impact de la sortie d'un neurone sur un autre. Au cours de l'apprentissage, ces poids sont ajustés pour minimiser l'erreur du réseau.

Les neurones sont organisés en couches qui peuvent être classées en fonction de leur position dans le réseau : entrée, cachée et sortie. La couche d'entrée reçoit les données, et la couche de sortie donne le résultat final de tous les traitements de données effectués par le réseau de neurones artificiels. Un réseau peut contenir une ou plusieurs couches cachées, ce qui détermine s'il s'agit d'un réseau profond ou non. Les couches cachées

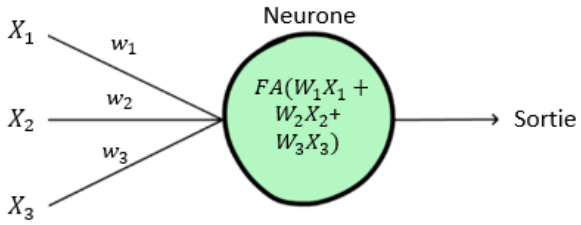


FIGURE 2.1 – Représentation du neurone.

traitent et analysent les données reçues soit de la couche d'entrée, soit des couches cachées précédentes.

Le processus de déplacement des données dans le réseau neuronal, de l'entrée à la sortie, est connu sous le nom de "feedforward". Par contre, le calcul du gradient de l'erreur du réseau par rapport à ses poids et l'ajustement des poids dans la direction opposée pour minimiser l'erreur est connu sous le nom de rétropropagation. Pendant son apprentissage, le réseau a pour objectif de minimiser la fonction de perte qui mesure la différence entre la sortie du réseau et les valeurs cibles réelles.

2.3 Réseaux génératifs antagonistes - pix2pix

Les réseaux génératifs antagonistes (Generative Adversarial Networks GAN), sont un des d'algorithmes d'apprentissage profond utilisés pour la génération de données synthétiques réalistes. Ils ont été introduits en 2014 par Goodfellow et son équipe et ils ont gagné beaucoup de popularité [85].

Un GAN est composé de deux réseaux de neurones, le générateur G et le discriminateur D. Le générateur G capte la distribution de probabilité des données réelles x, il prend en entrée un vecteur de bruit aléatoire z et essaie de générer des données $G(z)$ semblable et dans le même domaine que l'ensemble de données d'apprentissage réelles x (Voir la figure 2.2). Le discriminateur D prend en entrée des données provenant de deux sources différentes : l'ensemble de données d'apprentissage réelles x et les données générées par le générateur $G(z)$. Il estime la probabilité qu'un échantillon provienne des données d'apprentissage ou des données générées par G, classant ainsi les données comme des données réelles ou des données générées. Le discriminateur produit un scalaire unique $[0; 1]$, en utilisant x comme données positives ($D(x) = 1$) et en utilisant $G(z)$ comme

données négatives ($D(G(z)) = 0$) au cours de l'apprentissage.

Par conséquent, l'objectif du générateur est de générer des données réelles pour tromper le discriminateur, de manière à minimiser $log(1 - D(G(Z)))$, tandis que l'objectif du discriminateur est de classer correctement les données, de manière à maximiser $log(1 - D(G(Z)))$ et à maximiser $logD(x)$. Ainsi, les deux réseaux sont en compétition l'un contre l'autre, ils sont antagonistes au sens de la théorie des jeux et participent à un jeu à somme nulle, et la fonction objective du GAN original peut être définie comme suit :

$$\min_G \max_D V_{GAN}(D, G) = E_x[logD(x)] + E_z[log(1 - D(G(z)))], \tag{2.1}$$

où E_x est la valeur attendue lorsque x provient d'une distribution de données réelles, et E_z est la valeur attendue lorsque z provient d'une distribution aléatoire.

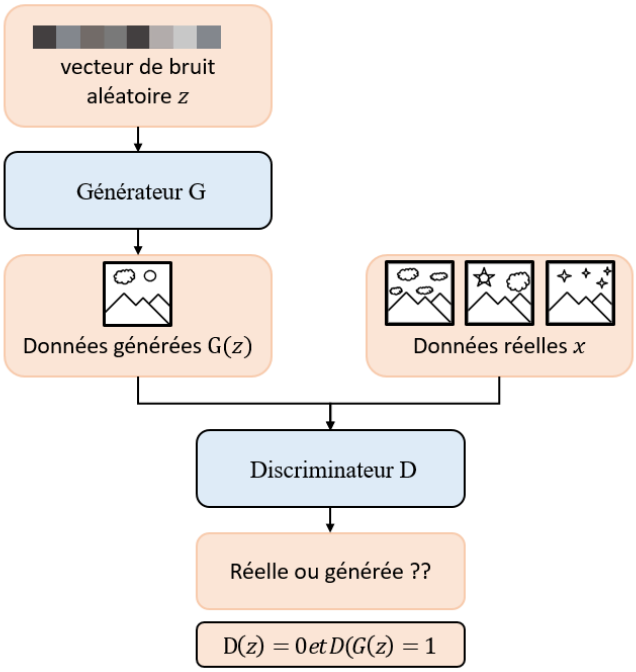


FIGURE 2.2 – Le schéma représentatif du GAN original.

De nombreuses versions de GAN ont été développées et modifiées, et leurs applications sont très diverses. Les GAN ont développé et introduit de nouvelles perspectives dans le domaine de l'intelligence artificielle, telles que la super résolution [87] [88], la translation d'image à image [89] [90] [91] [92], et la translation de texte à image [93] [94] [95], en raison de leur capacité à générer de nouvelles données. Les premiers GAN ont été utilisés pour

générer de nouvelles données réalistes sans qu'il soit possible de définir les caractéristiques des données à générer. Ensuite, il est devenu possible de préciser les caractéristiques des données générées, à l'aide d'une condition [96].

D'autres versions ont également été développées pour traiter le problème de la génération spécifique en utilisant le terme de classification. Le classifieur auxiliaire GAN (Auxiliary Classifier Generative Adversarial Network : ACGAN) a été développé pour améliorer l'entraînement du générateur, et permet de générer des données tout en contrôlant la classe souhaitée [97]. Le générateur prend en entrée le vecteur z avec une condition qui est la classe de l'image à générer, et le discriminateur tend à distinguer l'image réelle de l'image générée et à la classer. La fonction de perte du ACGAN se compose de la perte antagoniste et de la perte de classification auxiliaire. La classification fournit une supervision supplémentaire, stabilisant ainsi le processus d'apprentissage du GAN.

Le réseau génératif antagoniste avec classificateur auxiliaire polyvalent (Versatile Auxiliary Classifier Generative Adversarial Network VACGAN) introduit un réseau de classification en parallèle avec le discriminateur [98]. Le classifieur prend les données du générateur en entrée et cherche à les classer. L'erreur de classification est rétro-propagée à travers le classificateur et le générateur. Classiquement, le discriminateur tente à son tour de distinguer les données générées des données réelles.

La classification a pu être réalisée à l'aide de GAN, suite à quelques modifications et/ou compléments apportés au niveau du discriminateur. La classification a été utilisée pour forcer le générateur à générer une classe spécifique de données pour une entrée donnée, et pour améliorer le processus d'entraînement, afin de générer des données plus réalistes. Cette classification n'a pas été utilisée pour rendre le modèle un outil de classification.

L'une des versions qui nous intéresse dans la suite est le pix2pix qui est utilisée pour la transformation d'images entre différents domaines, la translation d'image à image.

pix2pix

Le pix2pix [89] est une version du GAN est conditionné par y lors de l'apprentissage. La condition y est une image, dont le but du générateur est de la régénérer à partir de l'image d'entrée x. Le générateur apprend la correspondance entre l'image observée x et le bruit aléatoire z, d'une part, et la condition y, d'autre part. Il essaie de produire en sortie $G(x, z)$ indistinguable de y. Le discriminateur D cherche à reconnaître l'image générée entre $G(x, z)$ et y. Le générateur et le discriminateur prennent en entrée l'image observée x et la condition y, comme le montre la figure 2.3.

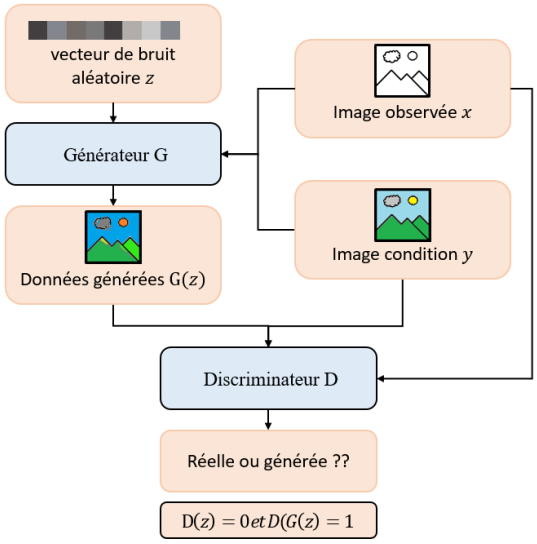


FIGURE 2.3 – Le schéma représentatif du pix2pix.

La fonction objective finale de pix2pix devient :

$$\min_G \max_D V_{pix2pix}(D,G) = E_{x,y}[logD(x,y)] + E_{x,z}[log(1 - D(G(x,z)))] + \lambda \mathcal{L}_{L1}(G) \quad (2.2)$$

où $E_{x,y}$ est la valeur attendue lorsque x provient d'une distribution de données réelles et y la condition, $E_{x,z}$ est la valeur attendue lorsque x provient d'une distribution de données réelles et z provient d'une distribution aléatoire, λ est un coefficient de réduction et $\mathcal{L}_{L1}(G)$ est la distance L1 entre y et $G(x,z)$:

$$\mathcal{L}_{L1}(G) = E_{x,y,z}[||y - G(x,z)||_1]. \quad (2.3)$$

où $E_{x,y,z}$ est la valeur attendue lorsque x provient d'une distribution de données réelles, y la condition, et z provient d'une distribution aléatoire.

Une fois l'apprentissage terminé, le discriminateur peut être supprimé et seul le générateur est utilisé (Voir la figure 2.4). Le générateur prend une image d'entrée x et tente de produire une image $G(z)$ basée sur x et dans le même domaine que celui dans lequel il a été entraîné.

Dans cette thèse, le pix2pix sera utilisé, et son générateur est responsable de la génération des données sous une condition qui sera spécifiée dans la section 2.5.

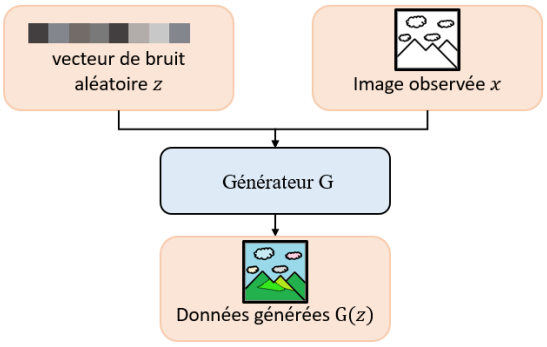


FIGURE 2.4 – Utilisation du pix2pix après entraînement.

2.4 Auto-encodeur

Les auto-encodeurs (AE) sont des réseaux de neurones, introduits pour la première fois en 1986, entraînés à reconstruire la donnée d'entrée en réduisant sa dimension [99]. L'AE se compose de deux éléments essentiels : l'encodeur E et le décodeur D. L'encodeur transforme l'entrée x en une représentation dans un espace de dimension plus faible appelé espace latent. La sortie de l'encodeur $E(x)$ est une représentation latente, utilisée comme entrée du décodeur. A partir de cette représentation latente, le décodeur tente a reconstruire l'entrée initiale x (Voir la figure 2.5).

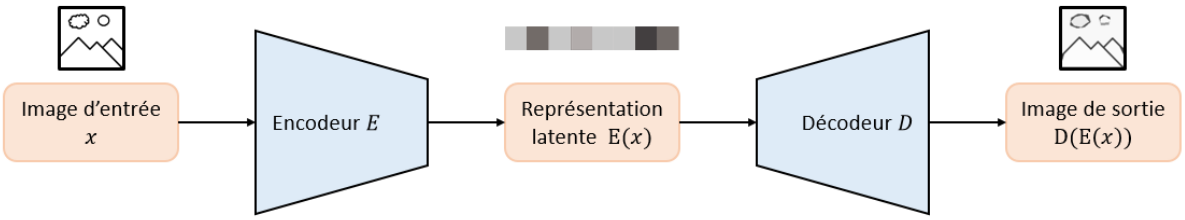


FIGURE 2.5 – Le schéma représentatif de l'AE.

Le processus d'apprentissage de l'AE consiste à minimiser la différence entre l'entrée initiale x et la sortie reconstruite par le décodeur, d'où la fonction objective de l'auto-encoder est :

$$arg\ min_{E,D} \quad E_x\ [\Delta(x,\ D(E(x))\ )] \tag{2.4}$$

où E_x est la valeur attendue lorsque x provient de la distribution des données d'entraîne-

ment, E et D sont l'encodeur et le décodeur respectivement.

Les AE sont principalement utilisés pour compresser et retirer les données utiles [100] [101]. Plusieurs modèles et versions d'auto-encodeurs ont ensuite été développés pour répondre à différents défis et tâches :

- Les auto-encodeurs de dé-bruitage ont été utilisé pour éliminer le bruit des données[102]. Ces AE sont entraînés sur des données perturbées par un bruit, et ont pour objectif de régénérer les données initiales dépourvues de bruit.
- Les auto-encodeurs sparce introduisent une contrainte de parcimonie pour activer uniquement un petit nombre de neurones dans la couche cachée [103]. Selon cette contrainte, seul un certain pourcentage de nœuds peut être actif dans une couche cachée. Les AE sparce réduisent le risque de sur-apprentissage et améliorent ainsi la généralisation du modèle.
- Les auto-encodeurs contractifs [104] sont plus résistants et plus robustes face à certaines perturbations d'entrée, alors que l'extraction de caractéristiques est moins sensible aux petites perturbations. Cette robustesse est obtenue en appliquant une pénalisation à la fonction de perte.
- Les auto-encodeurs variationnels (VAE) sont utilisés pour générer de nouvelles données à partir d'une distribution latente apprise [86]. La principale différence des VAE réside dans le fait que la sortie de l'encodeur est une distribution de probabilité. La fonction de perte des VAE est composée d'un terme de reconstruction, qui tend à rendre le schéma de codage-décodage aussi efficace que possible, et d'un terme de régularisation, qui tend à régulariser l'organisation de l'espace latent en faisant en sorte que les distributions renvoyées par l'encodeur soient proches d'une distribution normale standard. La régularisation est exprimée par la divergence de Kulback-Leibler entre la distribution retournée et une distribution gaussienne standard.

Les auto-encodeurs ont été utilisés dans une grande variété d'applications, et différents types ont été modifiés et combinés pour former de nouveaux modèles destinés à de nouvelles applications. La première application des AE a été la réduction des dimensions. Les auto-encodeurs ont également été largement utilisés pour les tâches de classification. L'encodeur peut être considéré comme un extracteur de caractéristiques dont la sortie est utilisée pour former un classificateur tel que le SVM. [105]. Avec le développement des VAE, les auto-encodeurs ont été utilisés dans des applications de génération de nouvelles données [106] et sont considérés comme des modèles génératifs.

Dans cette thèse, la partie encodeur de l'auto-encodeur sera utilisée seule, sans la partie décodeur, ainsi sa sortie est la représentation latente. Le vecteur latent généré par l'encodeur est considéré comme la donnée générée selon une condition qui sera expliquée dans la section suivante, et l'encodeur est considéré comme un générateur.

2.5 Classification par génération

Les méthodes et les modèles permettant de générer des données ont été largement développés et ont fait l'objet d'une attention particulière. Leur capacité à générer des données réalistes présente un grand intérêt. Ces méthodes ont permis la création de nouveaux domaines et de nouveaux défis. Les GAN et les AE ont été testés dans plusieurs applications de génération et ont montré de très bonnes performances. Ils ont également été utilisés pour des tâches de classification telles que l'ACGAN, le VACGAN et l'extraction de caractéristiques pour la classification à l'aide de l'AE. Mais cette classification ne nous dit rien sur la capacité de ces méthodes à différencier les différentes classes. Pour les GAN, la classification est un outil d'amélioration de performance, et se fait soit par l'ajout d'un classifieur, soit par le discriminateur qui est déjà un classifieur. Les auto-encodeurs permettent de préparer l'information nécessaire utilisée par un autre classifieur.

Dans cette thèse une méthode de classification par génération est développée inspirée par la méthode de classification à faible coût à l'aide d'auto-encodeurs [107]. Cette méthode présente un test réel de la capacité des méthodes générant des données à effectuer une classification. Dans cette méthodes, les modèles ont pour but de générer des données spécifiques en fonction de la classe de l'image d'entrée. Les données utilisées sont des images.

Dans ce qui suit, la création de la base de références nécessaire à la méthode de classification par génération sera expliquée, puis les réseaux utilisés seront présentées, ainsi que la méthode de comparaison et la prise de décision.

2.5.1 Base de références

La préparation de la base de références est la première étape de la classification par génération. Toutes les classes doivent être présentées par une donnée dans la base de références qui est considérée comme la référence de cette classe.

Soit X l'ensemble d'apprentissage contenant p classes et m images par classe. $X =$

$\{x_1^1, ..., x_1^m, ..., x_p^1, ..., x_p^m\}$. Pour chaque classe i, une image de référence r_i est choisie, où $0 < i \leq p$, comme le montre la figure 2.6. Un ensemble de références $R = \{r_1, ..., r_p\}$ est donc créé. Par conséquent, à chaque image x de l'ensemble d'apprentissage X correspond une référence r dans la base de références.

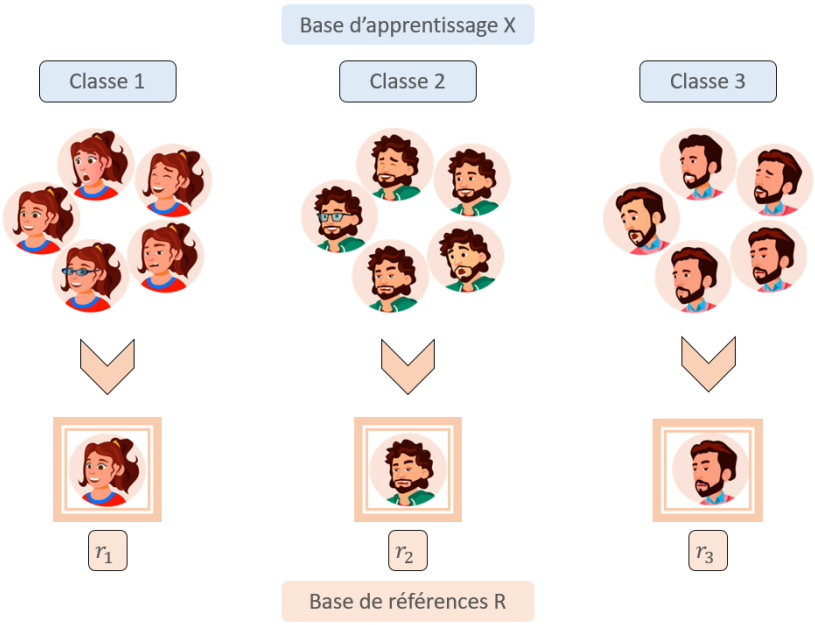


FIGURE 2.6 – Création de la base de références.

La donnée représentant la classe dans la base de références peut être choisie à partir de la base d'apprentissage de telle sorte que $r_i \in \{x_i^1, ..., x_i^m\}$, si possible. Cette donnée peut également être une donnée quelconque sans aucun rapport avec la classe qu'elle représente.

2.5.2 Les réseaux

Les réseaux utilisés dans cette méthode de classification sont appelés générateurs parce qu'ils génèrent les données importantes, même s'ils ne sont pas considérés comme des modèles génératifs. Pour classer des images à l'aide de la méthode de classification par génération, le réseau utilisé doit être capable de générer une sortie de même nature que les données dans la base de références. La sortie du réseau n'est pas une probabilité mais une donnée générée.

Les réseaux sont entraînés de manière à ce que, quelle que soit l'image d'entrée ap-

partenant à une classe i, ils génèrent en sortie l'image r_i représentant cette classe dans la base référence, comme le montre la figure 2.7. Par conséquent, les images de la base de référence peuvent être considérées comme les données représentatives de toutes les classes, ainsi que comme les données souhaitées obtenues par le générateur.

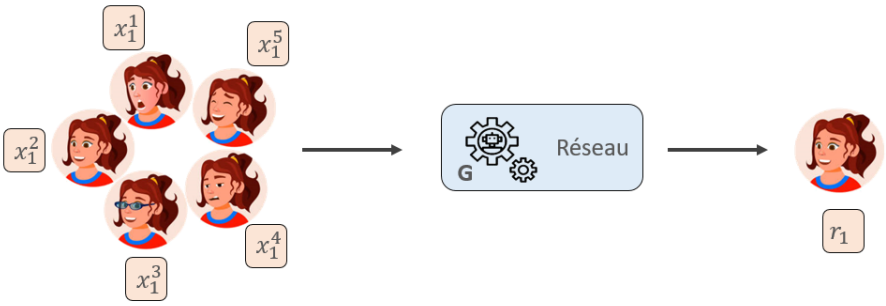


FIGURE 2.7 – Schéma de la classification par génération.

Le pix2pix et l'auto-encodeur ont été utilisés dans la classification par génération en tant que réseaux. Ils sont implémentés en utilisant la librairie Pytorch. La taille du lot (batch size) est fixée à 32 et le taux d'apprentissage (learning rate) à 0,0002. Les réseaux sont entraînés sur 100 époques.

2.5.3 Le pix2pix

Le pix2pix formé de deux réseaux, le générateur G et le discriminateur D peut être utilisé pour la classification par génération [108]. Pendant la phase d'apprentissage, le générateur G prend en entrée l'image $x_i^q \in X$ qui doit être classée, la condition $y = r_i \in R$ étant la référence correspondante de x_i^q. Le générateur tente de régénérer la référence r_i, vue comme la sortie souhaitée. Comme d'habitude, le discriminateur cherche à reconnaître l'image générée entre $G(x_i^q)$ et r_i.

Une fois le réseau correctement entraîné, le discriminateur est inutile et peut être supprimé ; seul le générateur est utilisé pour classer les données. Étant donné une image d'entrée, le générateur doit générer la sortie la plus similaire à la référence correspondante à sa classe.

Le pix2pix utilisé est celui de [89]. Le générateur est basé sur le réseau U (U-net) [109], dont l'architecture est détaillée dans la figure 2.8. Le classifieur patchGAN [110] est utilisé comme discriminateur dont l'architecture est présentée dans la figure 2.9.

Pour toutes les figures d'architecture des réseaux (2.8, 2.9 et 2.11), les couches des réseaux de neurones sont définies dans la figure 2.10.

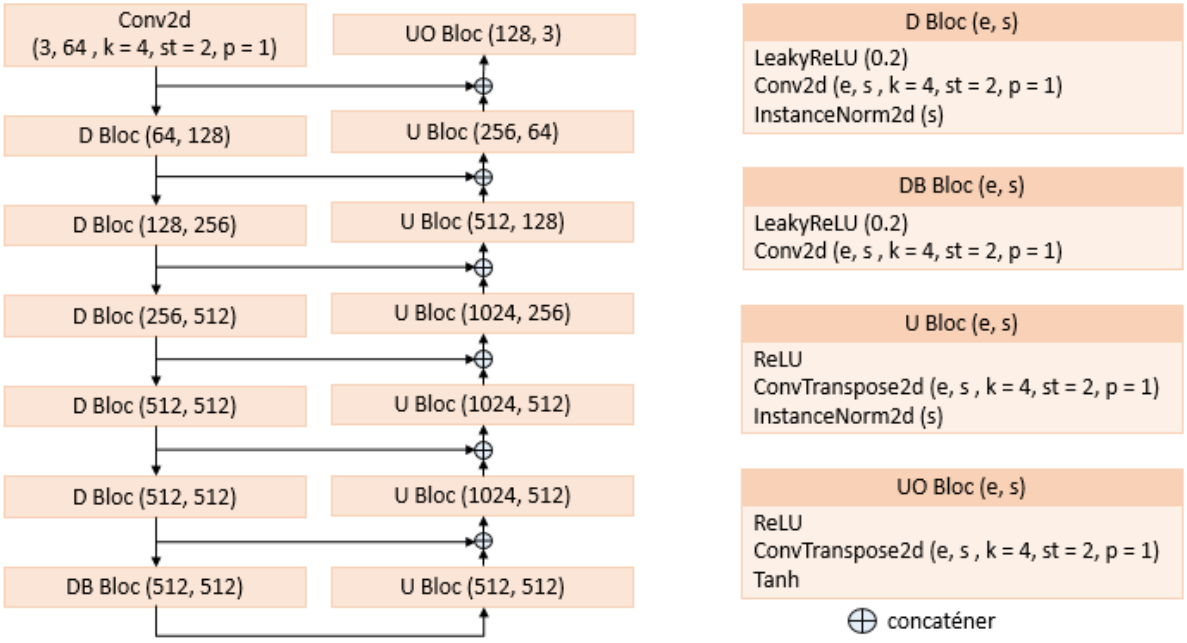


FIGURE 2.8 – L'architecture du générateur - U-net.

Le générateur prend en entrée une image couleur à trois canaux (RVB - Rouge, Vert, Bleu) de dimensions $128 \times 128 \times 3$. Il génère une image de sortie RVB de mêmes dimensions $128 \times 128 \times 3$. La base de référence doit donc comporter des images ayant les mêmes dimensions que l'image de sortie du générateur, c'est-à-dire des images de dimensions $128 \times 128 \times 3$.

2.5.4 L'auto-encodeur

L'auto-encodeur peut également être utilisé pour la classification par génération. Seule la partie encodeur est utilisée. Sa sortie, la représentation latente, est considérée comme la donnée générée importante. L'encodeur prend en entrée l'image $x_i^q \in X$ qui doit être classée, et est entraîné à régénérer la référence r_i correspondante à son entrée x_i^q.

Le ResNet-18 dont l'architecture est présentée dans la figure 2.8 est utilisé comme encodeur. Il prend en entrée une image couleur RVB de dimensions $128 \times 128 \times 3$. La représentation latente générée par l'encodeur est un vecteur latent de dimensions 1024.

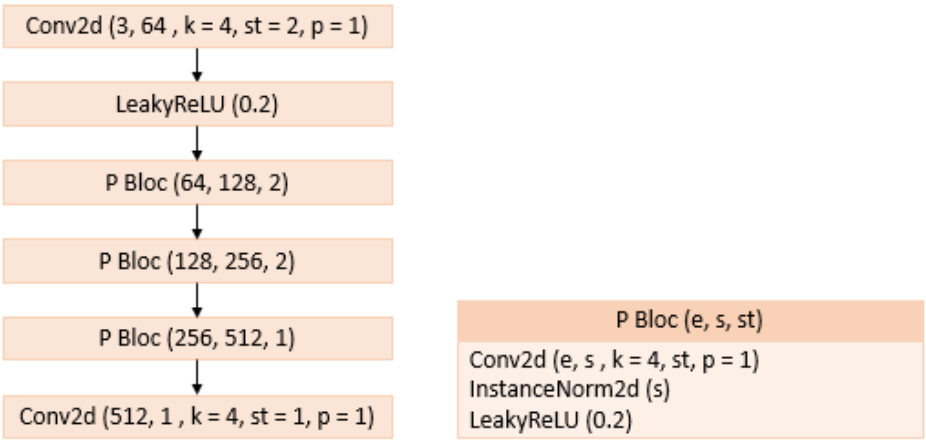


FIGURE 2.9 – L'architecture du discriminateur - PatchGAN.

La base de référence doit donc comporter des données de même nature que la sortie de l'encodeur, c'est-à-dire des vecteurs de dimensions 1024.

2.5.5 Comparaison et prise de décision

Le réseau est entraîné à générer en sortie la donnée de référence représentant la classe de l'image d'entrée. La donnée générée par le réseau $G(x)$ doit donc être comparée à toutes les données r_i de la base de références, afin de déterminer l'image de référence qui lui ressemble le plus. Cette comparaison se fait à l'aide de l'erreur quadratique moyenne (Mean Squared Error MSE) entre la donnée générée et les données de la base de références :

$$MSE_i = \frac{[r_i - G(x)]^2}{n \times m}, \tag{2.5}$$

Où $0 < i \leq p$, p le nombre de classes, n et m les dimensions de la donnée générée. Dans le cas de pix2pix $n = m = 128$, tandis que dans le cas de l'auto-encodeur $n = 1$ et $m = 1024$. L'image d'entrée est considérée comme appartenant à la classe j représentée par la donnée de référence ayant la valeur MSE la plus faible, donc la donnée la plus proche de la donnée générée.

$$MSE_j = \min_{0<i\leq p} \{MSE_i\}, \tag{2.6}$$

Dans l'application du contrôle d'accès, le choix "inconnu" est essentiel. Certaines personnes sont inconnues, n'appartiennent pas aux utilisateurs autorisés, ne font pas partie de la base d'entraînement, et doivent être rejetées par le système après avoir été consi-

Conv2d (i, o, k, st, p)

Couche de convolution 2D :
- i : nombre de canaux d'entrée
- o : nombre de canaux de sortie
- k : taille du noyau (kernel size)
- st : foulée (stride)
- p : rembourrage (padding)

BatchNorm2d (n)

Couche de normalisation par lots sur une entrée 4D:
- n : nombre de caractéristiques

PReLU

Couche de la fonction élémentaire :

$$PReLU(x) = max(0,x) + a * min(0,x)$$

MaxPool2d (k, st, p)

Couche de 2D max pooling :
- k : taille du noyau (kernel size)
- st : foulée (stride)
- p : rembourrage (padding)

AdaptiveAvgPool2d

Couche de regroupement des moyennes adaptatives en 2D

Linear(i, o)

Couche de transformation linéaire :
- i : nombre de canaux d'entrée
- o : nombre de canaux de sortie

PReLU

Couche de la fonction élémentaire :

$$Sigmoid(x) = \sigma(x) = \frac{1}{1 + \exp(-x)}$$

Flatten

Aplatit l'entrée en la transformant en un tenseur unidimensionnel

BatchNorm1d (n)

Couche de normalisation par lots sur une entrée 2D:
- n : nombre de caractéristiques

LeakyReLU (ns)

Couche de la fonction élémentaire :

$$LeakyReLU(x) = max(0,x) + ns * min(0,x)$$

- ns : pente négative

InstanceNorm2d

Couche de normalisation des instances sur une entrée 4D

ConvTranspose2d (i, o, k, st, p)

Couche de de convolution transposé 2D :
- i : nombre de canaux d'entrée
- o : nombre de canaux de sortie
- k : taille du noyau (kernel size)
- st : foulée (stride)
- p : rembourrage (padding)

Tanh

Couche de la fonction Tangente hyperbolique par élément

FIGURE 2.10 – Définition des couches des réseaux de neurones.

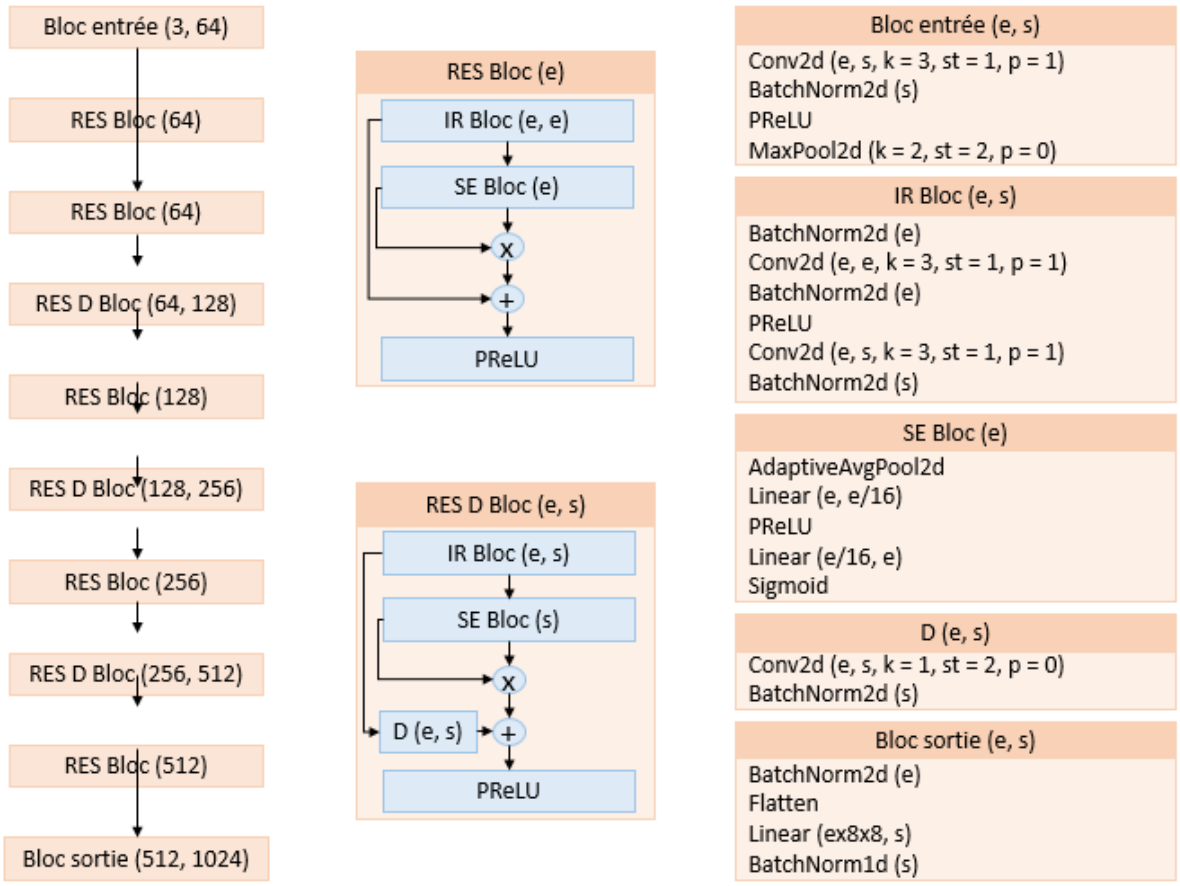


FIGURE 2.11 – L'architecture de l'encodeur - ResNet-18.

dérées comme inconnues. Dans cette approche, la personne est supposée inconnue si le réseau ne parvient pas à générer une donnée de sortie proche de l'une des données de la base de références. Dans ce cas, le réseau génère une sortie sans caractéristiques, une combinaison de différentes données de référence, une sortie plus proche d'un bruit.

Afin de décider si une entrée est connue ou non, la différence entre les deux plus petites valeurs MSE précédemment calculées doit être évaluée :

$$D = \frac{min_2 - min_1}{min_2} \tag{2.7}$$

Avec :

$$min_1 = MSE_j = \min_{0<i\leq p} \{MSE_i\} \tag{2.8}$$

$$min_2 = \min_{\substack{0<i\leq p; \\ i\neq j}} \{MSE_i\} \tag{2.9}$$

Si D est supérieur à un seuil prédéfini s, le réseau a proprement généré la sortie de telle sorte qu'elle est très proche de la référence correspondante et assez éloignée des autres références, l'entrée est donc considérée comme une personne connue appartenant à la classe j correspondant à min_1.

Dans le cas contraire, si D est inférieur au seuil, la sortie générée est douteuse, ressemblant à deux ou plusieurs références en même temps, ou très éloignée de toutes les références. Dans ce cas l'entrée est considérée comme une personne inconnue.

Le schéma final de la méthode de classification par génération devient alors comme le montre la figure suivante 2.12

2.6 Résultats

2.6.1 Base de données

La méthode de classification par génération a été testée, dans un premier temps, à l'aide de la base de données ORL (Olivetti Research Laboratory) [111]. Cette base, normalement utilisée pour la reconnaissance faciale, contient 400 images faciales frontales de 40 individus différents. Chaque classe est composée de 10 images prises dans des poses, des expressions faciales et des angles différents, avec ou sans lunettes. Un exemple de classe est présenté dans la figure 2.13a. Une méthode de détection des visages a été appliquée aux images de la base de données (Voir figure 2.13b) afin de sauvegarder et de traiter

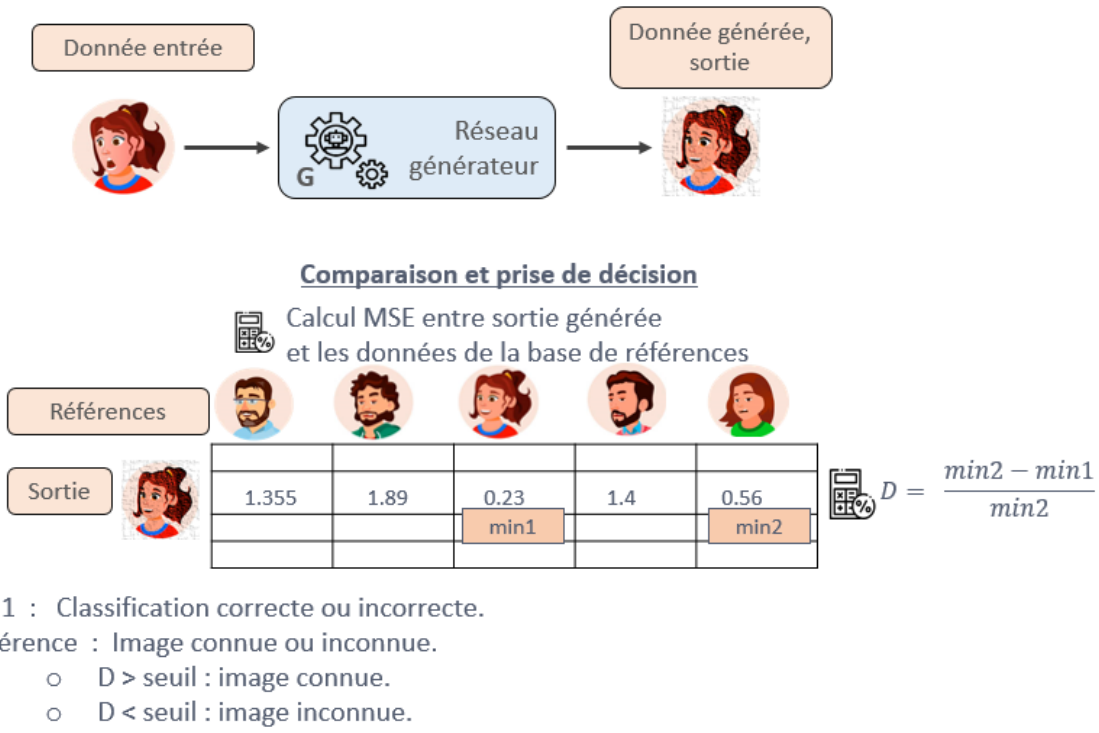
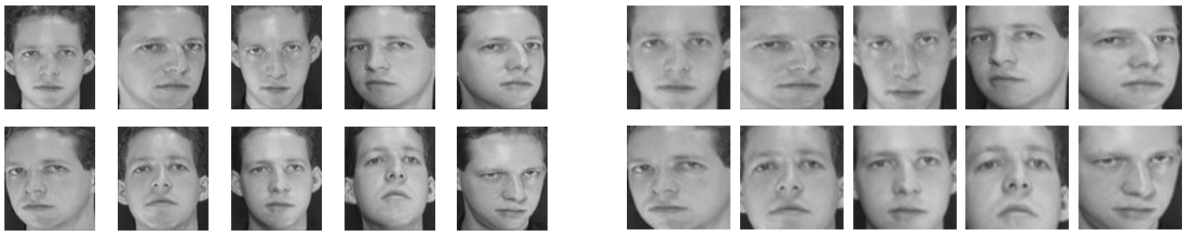


FIGURE 2.12 – schéma final de la méthode de classification par génération.

uniquement les visages.



(a) Images originales. (b) Images après détection du visage.

FIGURE 2.13 – Exemples d'images d'une classe de la base de données ORL.

Afin d'adapter la base de données ORL à l'application de contrôle d'accès, certaines classes e doivent être exclues de la base d'apprentissage, et leurs images constituent l'ensemble exclu E. Les personnes appartenant aux classes exclues sont considérées comme inconnues. Les images des autres classes de personnes connues sont divisées en deux ensembles : l'ensemble d'apprentissage A contenant a images par classe est utilisée pour l'apprentissage du réseau, et l'ensemble de test T contenant les t images restantes par classe n'appartenant pas à A est utilisée pour le test.

Une autre application, autre que le contrôle d'accès, a été testée en utilisant la base de données MNIST afin de vérifier les performances de la méthode de classification basée sur la génération. La base MNIST [112] contient 70 000 images de chiffres écrits à la main réparties en 10 classes, dont 60 000 images sont utilisées pour l'apprentissage et les 10 000 images restantes sont utilisées pour le test. La figure 2.14 montre quelques exemples des images dans MNIST.

Dans tout ce qui suit, la base de données ORL sera utilisée par défaut. En cas d'utilisation d'une autre base de données, une note sera clairement fournie.

FIGURE 2.14 – Exemple d'images de la base de données MNIST.

2.6.2 Matrice de confusion et évaluation

Pour définir correctement toutes les notions, le tableau 2.1 est utilisé. Une image appartenant à l'ensemble exclu E est toujours classée parmi les classes d'apprentissage connues. Cette image doit être considérée comme inconnue donc son D doit être inférieur au seuil e (cas ATN). Par contre, si D est supérieur au seuil, l'image est considérée comme étant une autre personne connue, ce qui est le pire des cas (cas AFP). Un autre cas indésirable est celui où une personne appartenant à une classe connue (appartenant à l'ensemble de test T) est mal classée et est considérée comme connue (cas FP). Dans le cas de classification erronée, la personne doit être considérée comme inconnue (cas TN). Une personne bien classée peut être considérée comme connue (cas TP) ou inconnue (non préférée - cas FN).

TABLE 2.1 – La matrice de confusion.

		Classification	
		Correcte	Incorrecte
Décision	Connue	Vrai positif TP	Faux positif FP / AFP
	Inconnue	Faux négatif FN	Vrai négatif TN / ATN

La méthode de classification par génération est évaluée à l'aide des paramètres suivants :

— Précision de la classification :

$$PC = \frac{TP + FN}{TP + FN + FP + TN}. \tag{2.10}$$

c'est le nombre d'images test de l'ensemble T qui sont bien classées indépendamment de la tâche de contrôle d'accès, donc indépendamment du seuil prédéfini.

— Précision du contrôle d'accès sur les images test de l'ensemble T uniquement :

$$PT = \frac{TP + TN}{TP + TN + FP + FN} \tag{2.11}$$

— Précision du contrôle d'accès sur les images exclues, les images de l'ensemble E seulement :

$$PE = \frac{ATN}{ATN + AFP} \tag{2.12}$$

— Précision du contrôle d'accès sur les images test et exclues, les images provenant

de la combinaison de T et E :

$$PS = \frac{TP + TN + ATN}{TP + TN + FP + FN + ATN + AFP} \tag{2.13}$$

Pour choisir le seuil e, la courbe ROC (receiver operating characteristic) a été utilisée. La courbe ROC est obtenue en traçant le taux de vrais positifs (TPR) et le taux de vrais négatifs (TNR) à différents seuils. Dans cette étude, les seuils sont compris entre 0 et 1, avec un pas de 0,01. Le seuil est choisi comme l'intersection des deux taux, TPR et TNR.

$$TPR = \frac{TP}{TP + FN} \tag{2.14}$$

$$TNR = \frac{TN + ATN}{FP + AFP + TN + ATN} \tag{2.15}$$

2.6.3 Choix de la référence

Les résultats générés par le réseau générateur-classificateur sont essentiellement basés sur les données de la base de références. La création de cette base est par conséquent une étape importante et essentielle dans la méthode de classification par génération. Comme indiqué dans la section 2.5.1 et selon le réseau générateur utilisé, les données de référence peuvent soit être choisies à partir de la base d'apprentissage, soit être des données quelconques aléatoires sans aucun lien avec la classe qu'elles représentent. Ainsi, le choix de ces références est un facteur intéressant qui doit être évalué et étudié dans ses diverses possibilités afin de mieux comprendre et de vérifier son effet sur les performances de la méthode de classification par génération.

Vu que la sortie du générateur pix2pix est une donnée de même nature (une image) et de même dimensions que la donnée d'entrée, la diversité des possibilités de création de la base de données de référence existe, contrairement au cas du réseau d'auto-encodeurs. La sortie de l'auto-encodeur étant un vecteur de dimension 1024, les données de la base de références doivent l'être également. Dans ce cas, la base de références ne peut donc pas être formée à partir de données appartenant à la base de données d'apprentissage qui sont des images de dimension $128 \times 128 \times 3$. Par conséquent, le réseau pix2pix sera utilisé pour étudier l'effet de la sélection des références et l'efficacité de la création d'une base de références aléatoire.

La base de données utilisée dans cette étude est la base ORL. Toutes ses classes sont utilisées pour l'apprentissage, l'ensemble exclu E est donc vide. Dans tous les tests,

l'ensemble d'apprentissage A est composé de 360 images au total, avec 9 images par classe ($a = 9$), et l'ensemble de test T est composé de 40 images au total, avec l'image restante d'indice imt de chaque classe ($t = 1$).

2.6.3.1 Références de la base d'apprentissage

Chaque classe de la base de données ORL se compose de 10 images différentes. Les différentes images de chaque classe ont été testées comme images de référence. Dans chaque test, l'image d'indice ref, appartenant à A, a été ajoutée à la base de référence, où $0 < ref \leq 10$, et $ref \neq imt$.

2.6.3.2 Références aléatoires

Pour chaque classe, une image aléatoire de référence de dimension $128 \times 128 \times 3$ est créée. L'image aléatoire est composée de 64 blocs de dimension 16×16 chacun, formés de 3 nombres entiers aléatoires représentant les trois canaux RVB. La figure 2.15 montre quelques exemples de références aléatoires (référence i) créées pour différentes classes i.

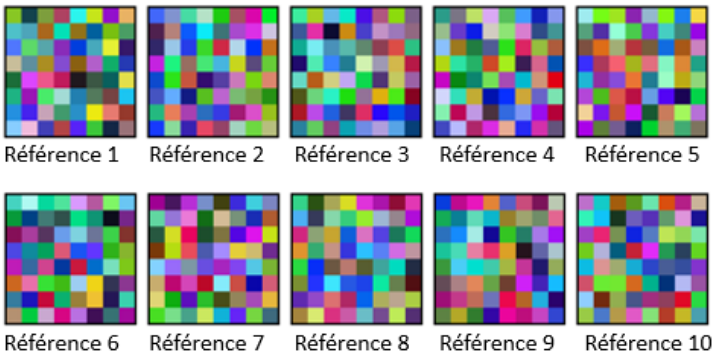


FIGURE 2.15 – Les références aléatoires des dix premières classes de la base de données ORL.

2.6.3.3 Résultats

Le tableau 2.2 montre les différentes précisions de classification PC obtenues en changeant l'image de référence choisie pour chaque cas. Dans ce tableau, ref désigne l'indice de l'image de référence choisie pour chaque classe, et imt désigne l'indice de l'image test utilisée. L'indice RA de ref correspond à la référence aléatoire.

TABLE 2.2 – La précision de classification *PC* dans les différents cas de la création de la base de références.

ref	1	2	3	4	5	6	7	8	9	10	RA
imt	2	1	2	2	2	2	2	2	2	2	2
PC (%)	100	100	100	100	100	100	100	100	97.5	100	100

La figure 2.16 montre quelques exemples de résultats obtenus (sorties obtenues ou générées) sur des images d'entrée de test, au dessus de références choisies (sorties souhaitées) pour différents cas d'images de référence.

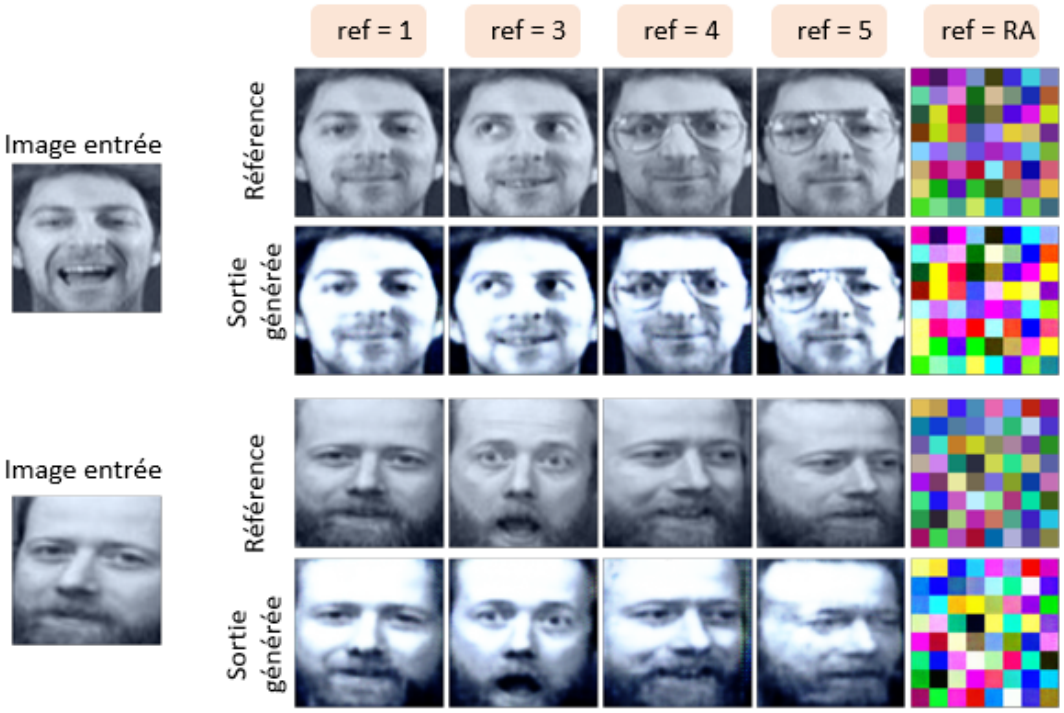


FIGURE 2.16 – Images générées par le pix2pix avec l'image test en entrée, au dessus des références pour différents cas.

Les résultats montrent que la performance de la classification est indépendante de la référence choisie. Quelle que soit la référence utilisée, la même performance est obtenue. Le réseau générateur a réussi à générer correctement la référence dans tous les cas, quelle que soit la relation entre l'image test d'entrée et la référence. Malgré les différences entre l'image test et l'image de référence dans les différents cas (expressions faciales diffé-

rentes, port de lunettes, sourire ou non, différentes inclinaisons du visage), le générateur a conservé les traits et les caractéristiques des visages. Dans le cas d'une référence aléatoire, le générateur a également réussi à générer une image similaire à l'image souhaitée, bien divisée en blocs, en préservant la séquence des couleurs. Il est clair que la luminosité de la sortie générée diffère de celle de la référence, mais cette différence n'affecte pas la classification des données.

La création de la base de référence peut donc se faire de n'importe quelle manière, sans donner d'importance aux références qui représenteront les classes, et sans affecter la performance de la classification. Toute donnée peut être utilisée comme référence, qu'elle appartienne ou non à la base d'entraînement et qu'elle soit ou non liée à la classe qu'elle représente. Par conséquent, le choix de la référence peut être fait de manière aléatoire, et la création d'une base de référence aléatoire est possible et efficace.

Dans ce qui suit, et dans le cas de références appartenant à la base d'apprentissage, l'image d'indice 1 de chaque classe est utilisée comme image de référence si elle n'appartient pas à l'ensemble de test T. Dans le cas contraire, l'image d'indice 2 ou 3 est utilisée comme référence.

2.6.4 Classification

Pour évaluer la performance de la méthode de classification par génération, les deux réseaux générateurs, le pix2pix et l'auto-encodeur, ont été utilisés sur différentes distributions (a, t) de la base de données ORL, où a est le nombre d'images par classe utilisé pour l'apprentissage, et t est le nombre d'images par classe utilisé pour le test. Dans l'évaluation de la classification, toutes les classes de la base de données sont utilisées pour l'apprentissage et l'ensemble exclu E est vide.

Dans le cas du générateur pix2pix, la base de référence peut être créée soit à partir de l'ensemble d'apprentissage, soit de manière aléatoire. Les deux possibilités ont été expérimentées.

La sortie du générateur d'auto-encodeur est un vecteur de dimension 1024. La base de référence dans ce cas est formée de vecteurs de dimensions 1024 avec des valeurs aléatoires d'une distribution normale (gaussienne). Pour visualiser les résultats, le vecteur est transformé en une image de dimensions $152 \times 144 \times 3$ composée de 1026 blocs de dimensions 8×8. Chaque bloc a une valeur du vecteur de référence, et les deux derniers blocs de l'image représentative ont des valeurs nulles. La figure 2.17 montre quelques exemples des images représentatives des vecteurs références crées pour différentes classes

i.

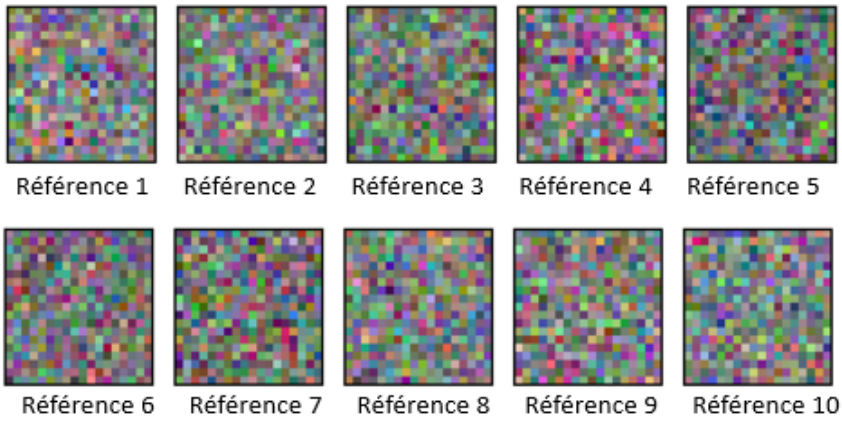


FIGURE 2.17 – Les images représentatives des références créées pour l'auto-encodeur pour les dix premières classes de la base de données ORL.

Dans cette section, toutes les précisions de classification (PC) correspondent à la valeur moyenne de 10 répétitions, en changeant à chaque répétition les images utilisées pour l'entraînement et le test des réseaux.

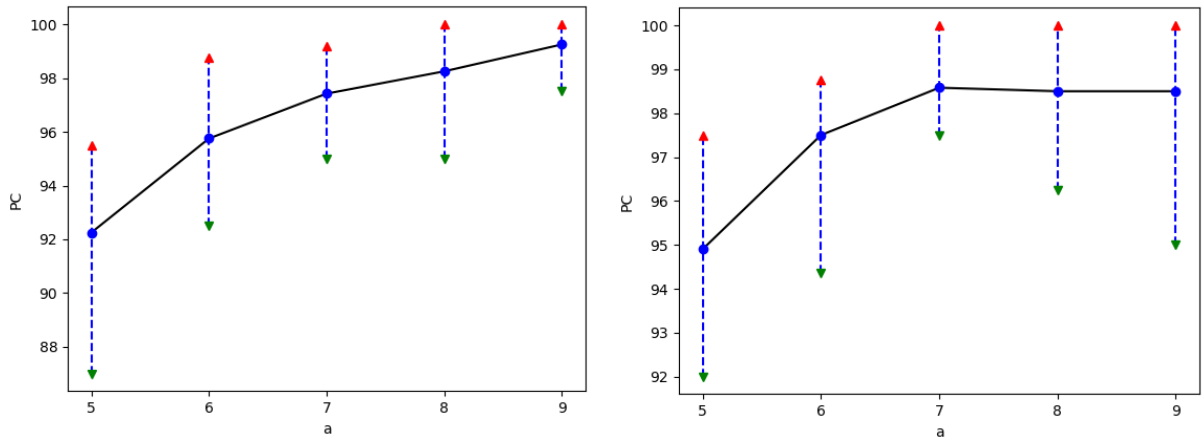
Le tableau 2.3 montre la précision de classification moyenne des 10 tests, ainsi que l'écart-type, pour différentes valeurs de (a, t) et pour les deux réseaux générateurs pix2pix et auto-encodeur. Dans le cas de pix2pix, les résultats pour les deux cas de références, appartenant à la base de données (références : B) et aléatoires (références : A), sont présentés.

TABLE 2.3 – Les précisions de classification (PC %) et les écarts types (±) pour les différents réseaux utilisés (pix2pix et auto-encodeur), pour différentes valeurs de (a, t), où B désigne les références appartenant à la base d'apprentissage et A désigne les références aléatoires.

Réseau	**Références**	**(a, t)**	(5, 5)	(6, 4)	(7, 3)	(8, 2)	(9, 1)
pix2pix	B	**PC**	92.25	95.75	97.42	98.25	99.25
		Écart-type	2.57	1.972	1.261	1.785	1.145
pix2pix	A	**PC**	94.9	97.5	98.58	98.5	98.5
		Écart-type)	1.758	1.281	0.917	1.658	2
AE	A	**PC**	95.65	97.59	99.08	98.87	99
		Écart-type	1.361	1.046	0.583	1.179	1.658

Dans tous les cas présentés dans le tableau, de bonnes précisions de classification sont obtenues. Avec seulement cinq images par classe utilisées pour entraîner le réseau, les performances obtenues peuvent être considérées comme satisfaisantes, et cette méthode peut donc être utilisée dans les cas où les données ne sont pas largement disponibles. Cependant, la précision de la classification augmente avec le nombre d'images par classe utilisées pour l'apprentissage, ce qui est logique et évident. Ainsi, si de meilleures performances sont requises, plus de données sont nécessaires. Dans tous les cas, les valeurs de l'écart-type sont faibles, ce qui confirme la grande reproductibilité et la fiabilité des résultats, reflétant la stabilité et la cohérence de cette méthode.

Les graphiques montrant l'évolution des valeurs de PC en fonction du nombre d'images d'apprentissage a par classe pour les deux cas de référence du pix2pix et pour l'auto-encodeur sont présentés dans les figures 2.18 et 2.19 respectivement. Dans ces graphiques, l'intervalle de PC est indiqué par les valeurs maximales (triangles rouges) et minimales (triangles verts) obtenues. Cet intervalle de valeurs de précision est principalement dû au choix de l'image test par classe. Dans la base ORL, certaines images sont significativement différentes des autres images de la même classe.



(a) Référence appartenant à la base de données (b) Référence aléatoire

FIGURE 2.18 – Précision de classification moyenne avec les valeurs maximales et minimales en utilisant le réseau pix2pix pour différentes valeurs de a.

La figure 2.20 montre quelques images générées par les deux générateurs pix2pix et AE. Cette figure montre l'image test en entrée du réseau, la référence ou la sortie souhaitée, et les images générées pour différentes valeurs de (a, t). Dans le cas de l'AE, les images présentées sont les images représentatives des vecteurs, par conséquent, dans ce cas, la séquence de couleurs est la seule caractéristique importante.

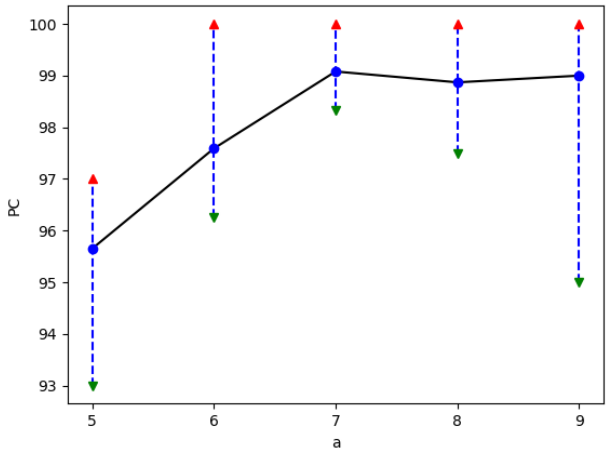


FIGURE 2.19 – Précision de classification moyenne avec les valeurs maximales et minimales en utilisant le réseau auto-encodeur pour différentes valeurs de a.

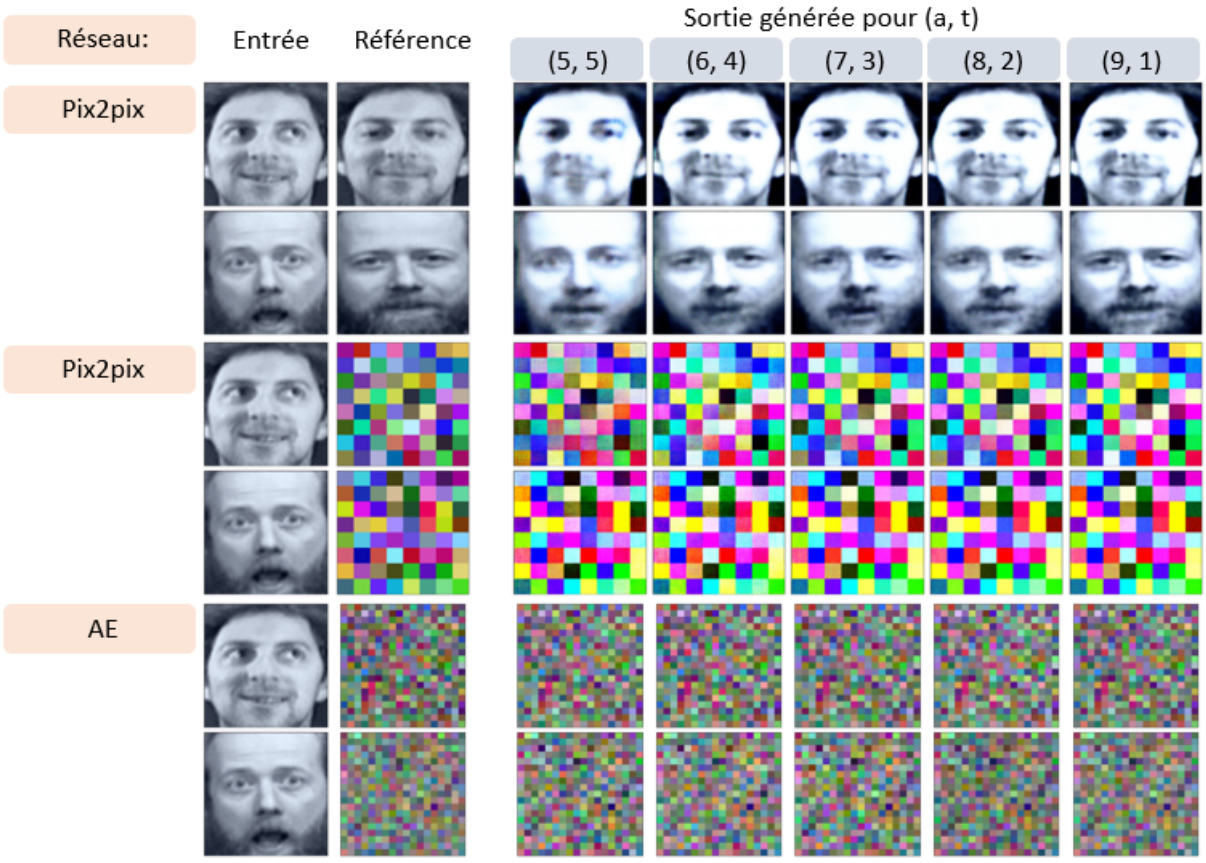


FIGURE 2.20 – Les résultats obtenus pour différentes valeurs (a, t) par les deux générateurs pix2pix et AE ayant comme entrée l'image test, ainsi que les références.

Il est remarquable qu'avec l'augmentation de a, la qualité de l'image générée dans le cas du générateur pix2pix s'améliore ; dans le cas d'une référence appartenant à l'ensemble d'apprentissage, l'image conserve ses caractéristiques et ses expressions, et dans le cas d'une référence aléatoire, les blocs sont de plus en plus bien bordés et ne se superposent pas. De même, à mesure que a augmente, et dans le cas d'une référence aléatoire pour les deux réseaux, la séquence de couleurs devient de plus en plus proche de celle de la référence. Cela explique mieux les résultats obtenus dans le tableau 2.3.

De plus, le tableau 2.3 montre que le cas des références aléatoires sur-performe le cas des références appartenant à la base de référence pour $a < 9$, pour le réseau pix2pix. Ceci peut s'expliquer par la mauvaise qualité de l'image du visage générée. Pour une petite valeur de a, l'image générée devient proche d'un visage flou et sans caractéristiques, car le réseau ne parvient pas à générer les caractéristiques correctement, ce qui ne permet pas de faire correspondre l'image générée à la référence correspondante. En outre, dans le cas de références aléatoires, même si la qualité de l'image générée est pauvre, le réseau arrive toujours à générer des blocs de couleur, avec à peu près la même séquence que la référence correspondante, bien qu'avec des degrés de couleur différents.

Le tableau 2.4 compare la méthode de classification par génération utilisant les différents cas avec d'autres algorithmes de détection de visages, PCA, FLDA, ICA, caractéristiques propres (eigen features) et visage propre (eigen face) [113]. La comparaison n'est pas très exacte, car il n'y a pas d'information sur le nombre d'images utilisées pour entraîner ces méthodes, ou si ces résultats sont la moyenne ou le meilleur. Pour la classification par génération, le résultat moyen obtenu avec la distribution des données$(a, t) = (9, 1)$ est considéré. Néanmoins, notre méthode de classification est plus performante que la plupart des autres méthodes. Si le meilleur résultat est pris en compte, notre méthode surpasse toutes les autres méthodes. En outre, si la précision de classification moyenne obtenue avec d'autres répartitions (a, t) est prise en compte, notre méthode présente des résultats acceptables par rapport aux autres méthodes.

2.6.5 Contrôle d'accès

Afin d'étudier la robustesse de la méthode de classification générative et son efficacité, une application de contrôle d'accès a été testée. Dans cette application, la capacité de la méthode de classification générative à distinguer les personnes connues et à rejeter les visages des personnes inconnues a été examinée.

Dans l'application de contrôle d'accès, l'ensemble exclu ne doit pas être vide ; il est

TABLE 2.4 – Précision de la classification de la méthode de classification par génération contre l'ACP, la FLDA, l'ICA, les caractéristiques propres et le visage propre sur la base de données ORL.

Méthode	PC (%)
PCA	94.30
FLDA	98
ICA	99.14
Caractéristiques propres	87
Visage propre	87.4
pix2pix (B)	99.25
pix2pix (A)	98.5
AE (A)	99

composé de ne classes considérées comme des classes inconnues. Différentes valeurs de ne ont été utilisées dans le test, $ne \in \{1, 3, 5, 10, 20\}$. Pour chaque valeur, l'apprentissage et le test ont été répétés 40 fois, en changeant à chaque fois les classes exclues e. L'ensemble d'apprentissage A est constitué de 9 images par classe ($a = 9$), et l'ensemble de test T est constitué de l'image restante d'indice $imt = 2$ ($t = 1$)

Dans la suite, toutes les valeurs des paramètres mesurant la performance correspondent aux valeurs moyennes des 40 répétitions.

Le tableau 2.5 présente les résultats obtenus pour différentes valeurs de ne, en utilisant les deux réseaux de génération pix2pix et AE. Dans ce tableau, B désigne le cas des références appartenant à l'ensemble d'apprentissage, et A désigne le cas des références aléatoires. Dans tous les cas, les réseaux ont été capables de classer tous les individus appartenant aux classes connues, indépendamment de la tâche de contrôle d'accès et du seuil à définir.

Une fois le seuil défini, certaines images de l'ensemble de test, bien classées, sont considérées comme inconnues, ce qui explique pourquoi PT est inférieur à PC. La décision inconnue peut être due à la mauvaise qualité de l'image générée dans le cas de pix2pix, de sorte qu'elle n'est pas suffisamment proche de la référence correspondante. Cette décision inconnue peut également être due à la ressemblance des données générées (dans les deux cas de réseau) avec plusieurs références, de sorte que les données générées

TABLE 2.5 – Les paramètres (PC, PT, PE, PS en %) mesurant les performances des réseaux générateurs, pix2pix et AE, pour différents nombres *ne* de classes exclues de la base d'apprentissage, et pour les seuils choisis.

Réseau	**Références**	**ne**	1	3	5	10	20
pix2pix	B	**PC**	100	100	100	100	100
		seuil	0.651	0.688	0.665	0.68	0.68
		PT	91.6	91.69	90.5	90.08	89.12
		PE	91.75	91.83	90.5	90.01	89.2
		PS	91.63	91.75	90.5	90.1	89.12
pix2pix	A	**PC**	100	100	100	100	100
		seuil	0.761	0.683	0.753	0.775	0.78
		PT	94.1	97.36	95	93.42	93.5
		PE	94	83.92	95.15	93.5	93.46
		PS	94.1	91.34	95.09	93.5	93.47
AE	A	**PC**	100	100	100	100	100
		seuil	0.418	0.41	0.417	0.457	0.511
		PT	95.6	96.9	96.3	95.3	95.8
		PE	95.5	96.92	96.2	95.4	95.82
		PS	95.56	96.9	96.8	95.38	95.82

ne sont pas vraiment considérées comme ressemblant à la référence correspondante, mais qu'il existe un doute entre plusieurs références. La méthode de classification par génération permet de déterminer les différents inconnus, comme le montrent les valeurs élevées de PE. Généralement, cette méthode de classification réussit à distinguer les personnes connues et à rejeter les personnes inconnues en même temps (voir le tableau des valeurs PS).

Le seuil est choisi comme expliqué dans la section 2.6.2. Pour chaque valeur de *ne*, un seuil unique est choisi pour les 40 répétitions, en traçant les taux moyens TPR et TNR comme indiqué dans la figure 2.21. Cette figure correspond au cas du générateur pix2pix, avec des références appartenant à la base d'apprentissage.

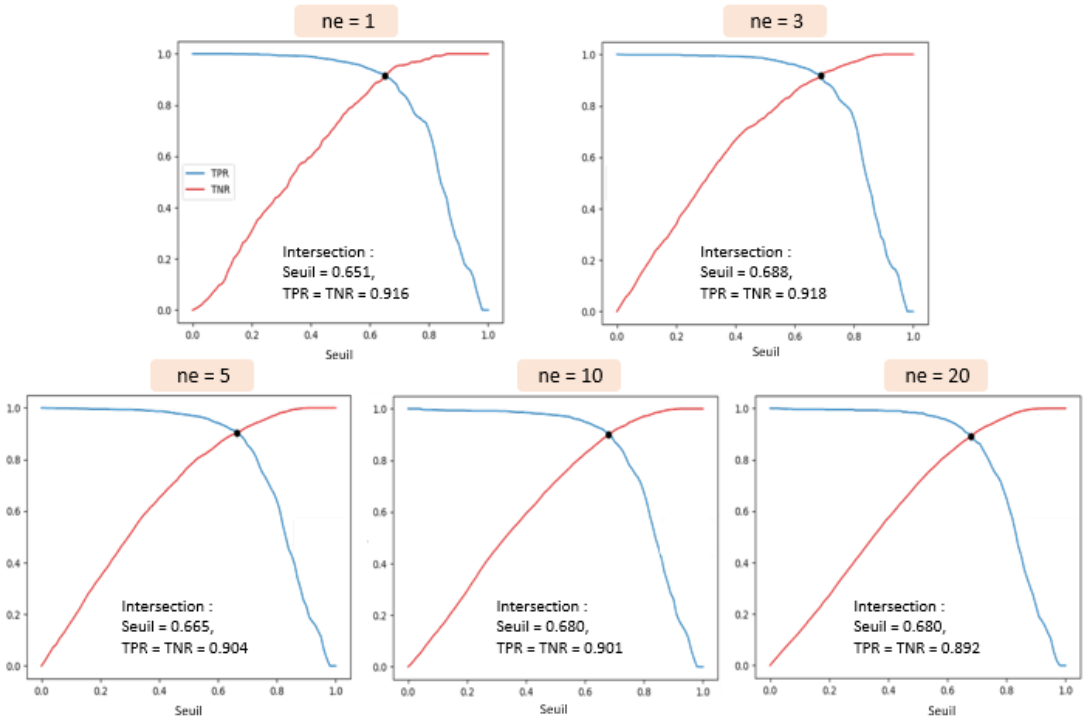


FIGURE 2.21 – Courbes ROC : TPR et TNR en fonction du seuil, pour différents nombres *ne* de classes exclues, dans le cas de pix2pix, avec des références de l'ensemble d'apprentissage.

La figure 2.22 montre la précision du contrôle d'accès sur les images test et les images exclues, pour différentes valeurs de seuil (PS), et met en évidence la valeur PS pour le seuil choisi, dans le cas du générateur pix2pix, avec des références appartenant à l'ensemble d'apprentissage. La valeur maximale de PS n'a pas été choisie (dans tous les cas de réseaux et de références), car il est nécessaire d'équilibrer PT et PE. Pour une valeur maximale de

SA, on obtient une valeur très faible de PT ou de PE. Dans ce cas, les réseaux ne seront pas efficaces pour la classification et le contrôle d'accès simultanément.

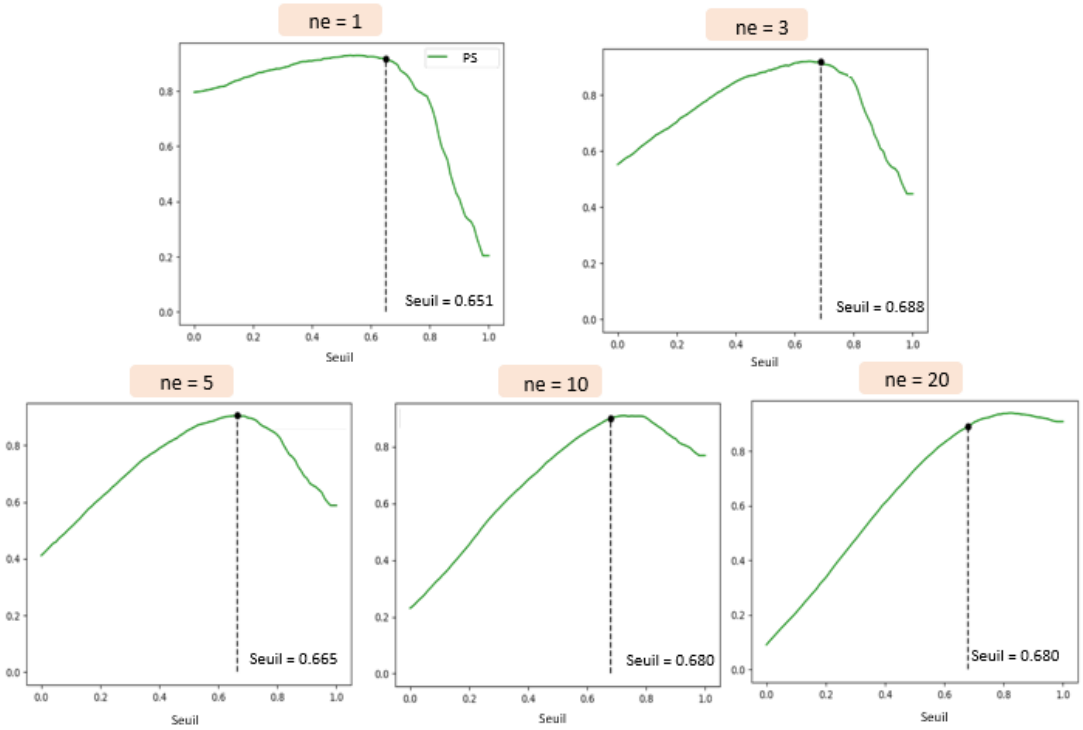


FIGURE 2.22 – Valeurs PS de pix2pix avec des références appartenant à l'ensemble d'apprentissage, pour différents nombres *ne* de classes exclues, en fonction du seuil.

La figure 2.23 montre quelques exemples d'images générées par pix2pix, et quelques exemples d'images représentatives de la sortie AE, pour différentes images d'entrée connues et inconnues, en face des références, dans le cas de $ne = 3$. Pour une personne connue, ayant une référence dans la base de références, la sortie générée est très similaire à la référence correspondante, et de bonne qualité (dans le cas de pix2pix). Dans le cas d'une personne sans images de référence, donc d'une personne inconnue, le résultat généré peut être considéré comme la combinaison de plusieurs références existantes. Dans le cas d'une référence appartenant à l'ensemble d'apprentissage, et pour une personne inconnue, le résultat généré est un visage flou, sans caractéristiques, ressemblant à plusieurs visages de la base de données de référence.

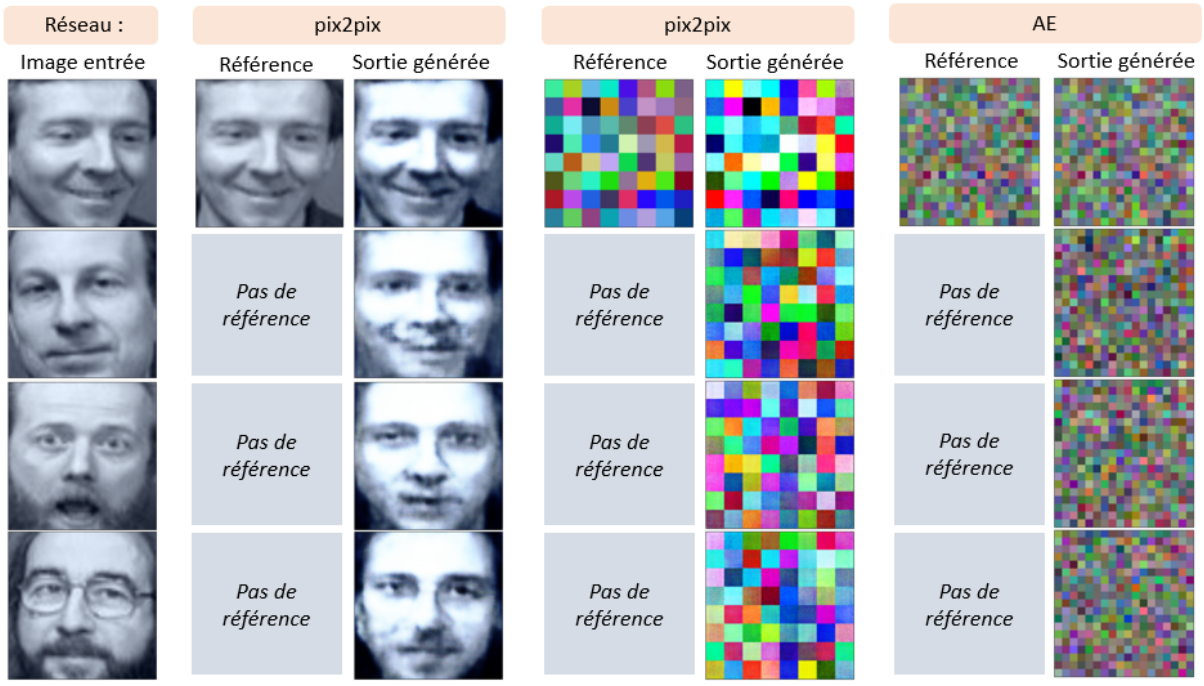


FIGURE 2.23 – Les résultats générés par les deux réseaux générateurs pix2pix et AE pour différentes images d'entrée connues et inconnues, dans le cas de $ne = 3$.

2.6.6 Reconnaissance des chiffres

Afin de tester la généralité de la méthode de classification par génération, l'application de reconnaissance de chiffres a également été testée. Dans cette application, la base de données MNIST a été utilisée. Dans cette application, seul le réseau pix2pix a été utilisé, avec des références appartenant à l'ensemble d'apprentissage. Le réseau a été entraîné sur 50 époques.

Les figures 2.24 et 2.25 montrent quelques exemples d'images générées par le générateur pix2pix dans les cas de bonne et de mauvaise classification respectivement, comparées aux références correspondantes. Le réseau a réussi à générer des références de bonne qualité. Le cas de mauvaise classification correspond souvent à des images difficiles à classer, telles que des images mal écrites.

Le tableau 2.6 compare les résultats de la méthode de classification par génération (pix2pix B) avec ceux de FSCGAN [114], d'un réseau neuronal convolutif (CNN) [107] et de LeNet-5 [115]. Pour FSCGAN, le cas de multiples fausses classes avec zéro échantillon non étiqueté est pris en compte afin d'être proche du cas étudié.

La méthode de classification par génération rivalise avec les méthodes précédemment

FIGURE 2.24 – Les résultats du MNIST correctement classés, générés par pix2pix, pour les images d'entrée, avec les références correspondantes.

FIGURE 2.25 – Les résultats du MNIST mal classés, générés par pix2pix, pour les images d'entrée, avec les références correspondantes.

TABLE 2.6 – Précision de la classification de la méthode de classification par génération (pix2pix B) par rapport à FSCGAN, CNN et LeNet-5 sur la base de données MNIST.

Méthode	PC (%)
pix2pix (B)	99.36
LeNet-5	99.05
CNN	98.62
FSCGAN	98.35

citées avec une légère différence de précision. La force de cette méthode par rapport à FSCGAN est que seul le générateur est utilisé pour effectuer la classification ; une fois que l'entraînement est terminé, le discriminateur peut être exclu. La comparaison de notre méthode avec LeNet et CNN montre l'efficacité de notre méthode en terme de classification, bien que LeNet et CNN soient développés pour la classification contrairement à pix2pix. Bien que la méthode de classification par génération n'atteigne pas l'état de l'art sur la base du jeu de données MNIST, il peut être noté que cette méthode est fortement compétitive par rapport à d'autres méthodes.

2.7 Conclusion

Ce chapitre présente une nouvelle méthode de classification basée sur la génération de données. Cette méthode est développée afin de tester la capacité réelle des réseaux générateurs de données (pix2pix et AE) à effectuer une vraie classification sans l'ajout de classificateurs. Dans cette méthode, le réseau générateur est entraîné à lier les données de chaque classe à une seule donnée que le réseau essaie de générer.

La méthode de classification par génération nécessite tout d'abord la création d'une base de références contenant toutes les données représentant toutes les classes. L'analyse des résultats obtenus sur différentes bases de références montre que les références peuvent être formées de manière aléatoire sans accorder beaucoup d'importance à leur choix.

La généralité de la nouvelle méthode de classification a été examinée en la testant sur différentes applications. Cette méthode a réussi à effectuer la reconnaissance faciale d'images de la base de données ORL et la reconnaissance de chiffres de la base de données MNIST. Les résultats obtenus en termes de précision de la classification sont satisfaisants

et capables de rivaliser avec les résultats d'autres méthodes de classification existantes.

En outre, la robustesse de la méthode de classification par génération a été testée en l'appliquant à la tâche de contrôle d'accès. Cette méthode a permis de reconnaître et de classer les différentes personnes connues et de découvrir les inconnues.

La méthode de classification par génération est implémentée en utilisant deux réseaux générateurs différents, le pix2pix et l'auto-encodeur. Dans les deux cas, de bonnes performances sont obtenues et le réseau générateur n'affecte pas fortement les résultats. La possibilité de généraliser les réseaux est donc presque approuvée. Pour appliquer cette méthode de classification, n'importe quel réseau capable de générer en sortie des données peut être utilisé.

Chapitre 3

Cryptage et tentative d'amélioration des résultats

3.1 Introduction

Préserver la sécurité des données partagées est un besoin qui existe depuis les civilisations anciennes, vu que certaines informations sont sensibles et privées, et ne devraient pas être accessibles à tout le monde lorsqu'elles sont partagées. L'histoire de la cryptographie a évolué depuis les simples méthodes de communication secrète jusqu'aux techniques cryptographiques complexes utilisées aujourd'hui dans le domaine de la sécurité numérique.

La première utilisation connue de la cryptographie a été l'utilisation de hiéroglyphes dans les tombes des souverains et des rois décédés [116]. Ces hiéroglyphes étaient conçus pour raconter la vie du roi et rappeler les hauts faits de sa vie d'une manière plus royale et plus importante, et non pour cacher le texte. Au fil du temps, ces écritures sont devenues de plus en plus compliquées et nécessitaient d'être déchiffrées pour être comprises.

L'une des méthodes de chiffrement simples les plus connues est le "chiffrement par décalage de César" de Julius Ceasar, utilisé initialement pour transmettre des messages secrets aux généraux de l'armée. La méthode consiste à décaler les lettres d'un message d'un nombre défini (qui représente la clé de cryptage), et le destinataire de ce message avance à son tour les lettres du même nombre pour obtenir le message original [117].

Différentes méthodes de cryptage et des techniques améliorées ont été développées. La méthode de substitution polyalphabétique [118] utilise plusieurs alphabets de substitution, en prenant deux disques de tailles différentes, l'alphabet clair est écrit sur la circonférence du disque inférieur et l'alphabet chiffré sur celle du disque supérieur. En déplaçant le disque intérieur d'une position à l'autre, différentes lettres du texte chiffré sont placées contre les lettres du texte clair, et chaque nouvelle position crée donc un nouvel alphabet chiffré.

Pendant la seconde guerre mondiale, la machine allemande Enigma, l'un des dispositifs de cryptage les plus connus, a été développée [119]. Elle utilise un système de cryptage mécanique à base de rotors pour coder les messages.

Aujourd'hui, le cryptage est un élément essentiel de la sécurité numérique, qui couvre les différentes formes de communication. Les protocoles cryptographiques sécurisent et protègent les échanges en ligne, les informations sensibles et permettent d'assurer la confidentialité des communications numériques. De nombreuses méthodes de cryptage ont été particulièrement développées pour sécuriser et protéger les images numériques. Ces méthodes permettent de préserver la confidentialité des images lors de leur transmission ou de leur stockage.

Les méthodes de cryptage d'images peuvent être divisées en deux catégories selon le domaine dans lequel le cryptage est appliqué : les méthodes spatiales et les méthodes fréquentielles [120]. Dans les méthodes du domaine spatial, les différentes procédures sont appliquées directement aux pixels qui composent les images. Les méthodes du domaine fréquentiel transforment les images du domaine spatial au domaine fréquentiel, où les différentes procédures de cryptage sont appliquées [121]. Ce type de cryptage présente l'avantage de pouvoir être implémenté optiquement.

D'autre part, les systèmes d'intelligence artificielle nécessitant la disponibilité des données sont utilisés dans divers domaines, dont certains utilisent des données personnelles privées. Cette dualité entre disponibilité et sécurité des données représente l'un des principaux défis auxquels est confrontée l'application de l'intelligence artificielle. Afin de préserver la sécurité des données, le développement de systèmes d'intelligence artificielle capables d'être entraînés sur des données cryptées et de les traiter, ainsi que le développement de méthodes de cryptage permettant de cacher les informations significatives aux humains et de conserver les informations indispensables pour les systèmes d'intelligence artificielle et les réseaux de neurones en particulier, sont devenus des projets intéressants.

Dans ce chapitre, une nouvelle méthode de cryptage d'images basée sur la transformée en cosinus discrète (DCT) est développée. Cette méthode de cryptage est utilisée sur les données privées de la méthode de classification par génération, afin de créer un système de classification préservant la confidentialité des données en les cryptant. L'objectif de ce système est de pouvoir utiliser immédiatement les données cryptées, sans nécessiter de modifications majeures au niveau du classificateur. La méthode de cryptage développée peut alors être considérée comme une méthode de cryptage perceptuelle.

L'utilisation directe de données cryptées par les réseaux de neurones peut affecter les

performances, car le cryptage change, modifie et cache les informations. Une tentative pour augmenter les performances a été testée dans ce chapitre sur les données cryptées, en utilisant le principe de l'apprentissage par transfert (Transfer Learning) par le biais d'un réglage fin (Fine Tuning - FT).

Ce chapitre est organisé comme suit : Après cette introduction, la méthode de cryptage basée sur la DCT est présentée, étudiée et testée avec la méthode de classification par génération dans les applications de classification et de contrôle d'accès, où la base de référence est constituée de références aléatoires créées comme indiqué dans 2.6.3.2 et 2.6.4, en utilisant le pix2pix et l'AE. Ensuite, la tentative d'amélioration des performances à l'aide de FT est détaillée et testée. La robustesse du système entier est ensuite étudiée. Enfin, une conclusion incluant les perspectives conclut ce chapitre.

3.2 Méthode de cryptage

Une nouvelle méthode de cryptage d'images dans le domaine fréquentiel est présentée dans cette section. Cette méthode de cryptage est basée sur la transformée en cosinus discrète (DCT), qui est principalement utilisée dans le traitement des signaux et la compression d'images [122]. La DCT transforme mathématiquement un signal ou une image du domaine spatial au domaine fréquentiel. Dans cette méthode, le cas de l'image (de dimensions $n \times m$) est considéré, en appliquant l'équation DCT 2D suivante :

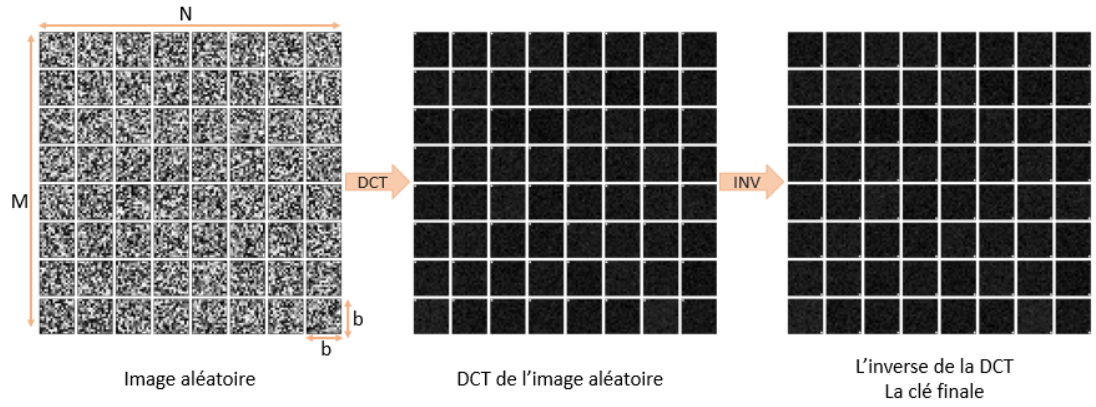
$$F(u,v) = \sqrt{\frac{2}{n}}\sqrt{\frac{2}{m}} \sum_{i=0}^{n-1} \sum_{j=0}^{m-1} \alpha(i)\alpha(j) cos\left[\frac{\pi u}{2n}(2i+1)\right] cos\left[\frac{\pi v}{2m}(2j+1)\right] f(i,j) \tag{3.1}$$

où $F(u,v)$ est le coefficient DCT de la ligne u et de la colonne v, $f(i,j)$ est l'intensité du pixel de la ligne i et de la colonne j de l'image, et :

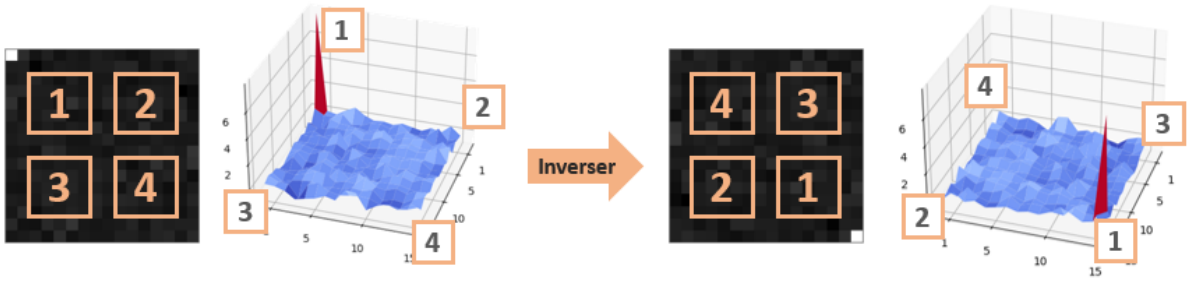
$$\alpha(x) = \begin{cases} \frac{1}{\sqrt{2}}, & \text{si } x = 0 \\ 1, & \text{sinon} \end{cases} \tag{3.2}$$

Le coefficient supérieur gauche, correspondant à $u = v = 0$, représente le courant continu ou le coefficient DC. Il représente la luminosité ou l'intensité moyenne de l'image. Le coefficient DC est généralement le plus grand coefficient et joue un rôle important dans la détermination de l'aspect général de l'image.

Pour se préparer à crypter les images, la clé de cryptage doit d'abord être créée. Une image aléatoire ayant les mêmes dimensions que l'image à crypter ($N \times M$) doit être créée et divisée en blocs de dimensions $b \times b$ comme le montre la figure 3.1a. La DCT est ensuite appliquée à chaque bloc ($n = m = b$ dans l'équation 3.1). Compte tenu de l'importance du coefficient DC du DCT et afin de réduire sa valeur pendant le cryptage, chaque bloc DCT est inversé de manière à ce que ce coefficient DC se retrouve dans le coin inférieur droit, comme le montre la figure 3.1b.



(a) Création de blocs aléatoires et application de la DCT.



(b) Inversion des blocs.

FIGURE 3.1 – Création de la clé de cryptage.

Pour crypter une image claire de dimensions $N \times M$, il faut d'abord la diviser en blocs de dimensions $b \times b$, puis appliquer la DCT à chaque bloc (voir figure 3.2). Ensuite, la DCT de chaque bloc est multipliée par le bloc correspondant de la clé de cryptage, ayant la même position. Comme le montre la figure 3.2, des pixels brillants apparaissent toujours dans les coins des blocs, ce qui facilite l'interprétation et permet de déduire l'application d'une méthode fréquentielle et de déterminer la taille b des blocs utilisés. Afin de casser cette forme de bloc, un mélange de pixels est appliqué sur la donnée cryptée entière (et non

au niveau du bloc) à l'aide d'une clé de cryptage de mélange. Ainsi, dans cette méthode de cryptage, deux clés sont utilisées : la clé des blocs et la clé de mélange.

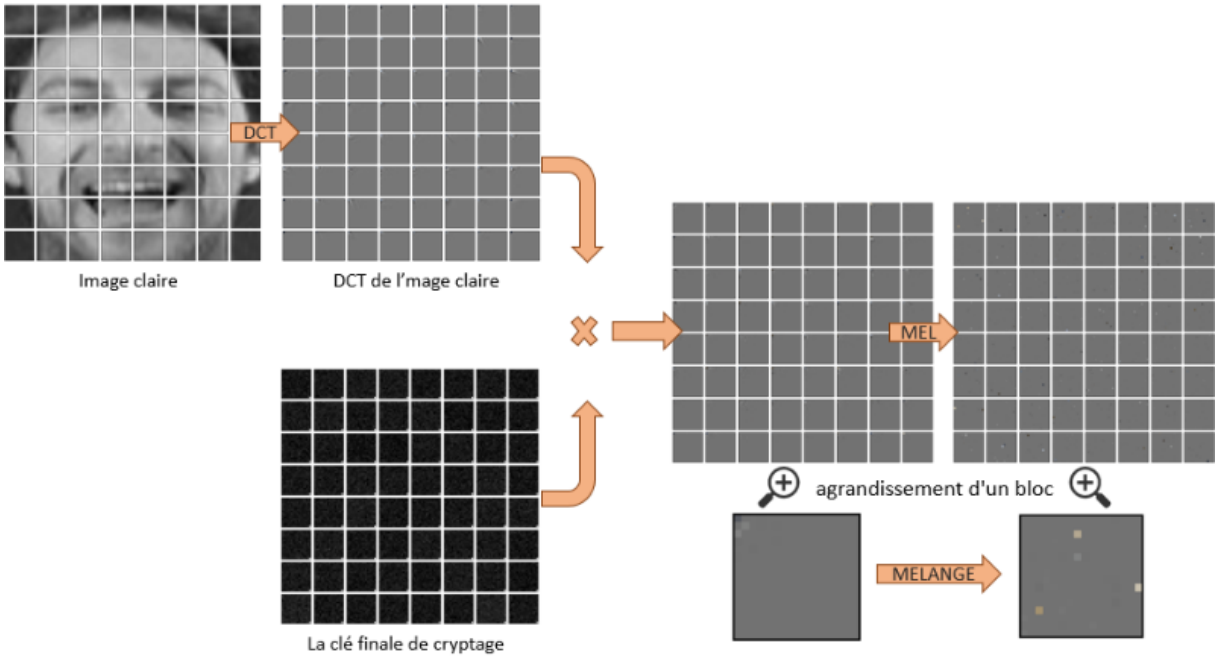


FIGURE 3.2 – La méthode de cryptage.

3.2.1 Choix de la taille des blocs

La taille de bloc b est un paramètre important sur lequel la méthode de cryptage est basée. La figure 3.3 montre les DCT des images à crypter pour différentes valeurs de la taille de bloc b utilisée. Il est clair qu'à mesure que b diminue, l'information obtenue devient de plus en plus claire. Pour $b = 1$, la DCT obtenue est la même que la donnée claire. Pour $b = 4$ et $b = 8$, les caractéristiques de l'image sont visibles, avec un peu d'attention on peut prédire que les données claires sont un visage dont les yeux et la bouche sont bien définis. Pour $b = 32$, l'information importante se trouve dans un petit nombre d'endroits, et l'application du cryptage peut provoquer la perte de cette information. Dans ce qui suit, la taille de bloc de cryptage $b = 16$ est considérée et utilisée. Dans ce cas, les informations importantes sont bien réparties et la forme des données initiales est cassée.

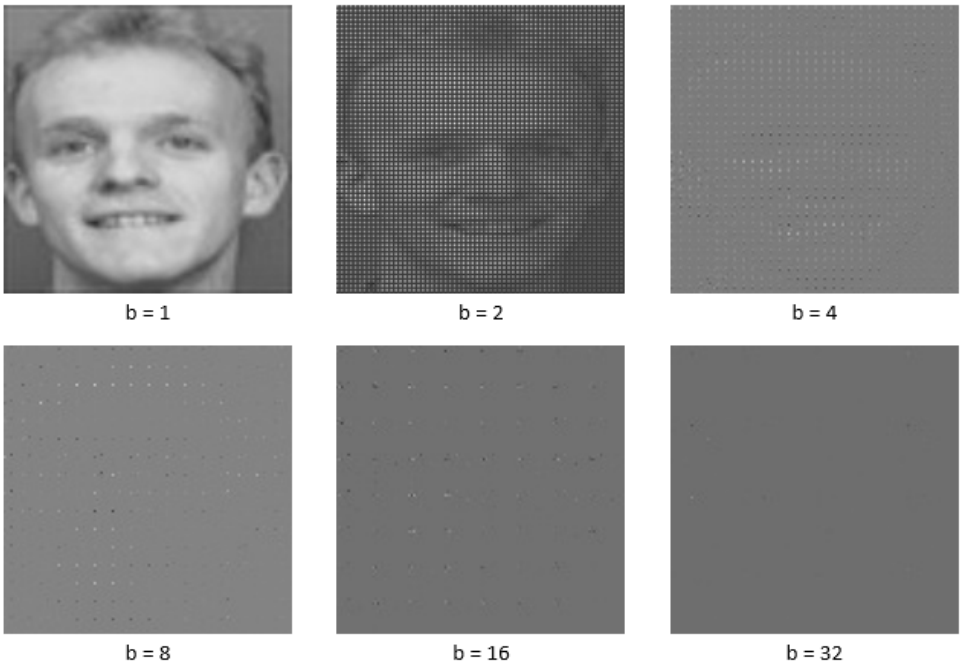


FIGURE 3.3 – DCT des images à crypter pour différentes valeurs de b.

3.2.2 Résultats

Afin de tester l'efficacité de la méthode de classification par génération sur des données cryptées et de déterminer l'influence de la nouvelle méthode de cryptage basée sur le DCT, la méthode de classification a été testée sur les données cryptées par la nouvelle méthode de cryptage, en choisissant des données aléatoires comme références, dans les deux applications de classification et de contrôle d'accès.

Le tableau 3.1 montre la précision de classification moyenne de 10 tests, où la distribution des images utilisées pour l'entraînement (a) et le test (t) des réseaux est modifiée à chaque test, ainsi que l'écart-type, pour les deux réseaux générateurs pix2pix et auto-encodeur.

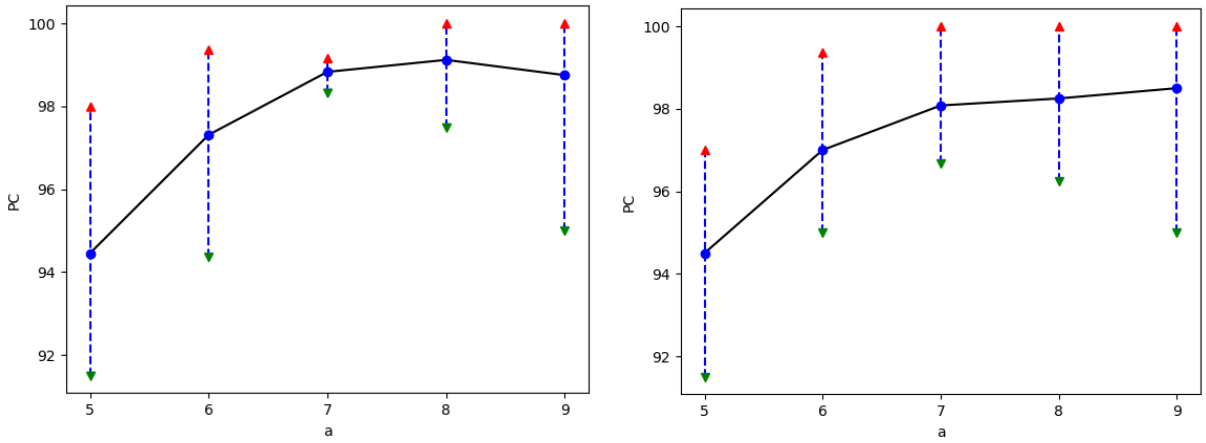
De bons résultats et de faibles valeurs d'écart type sont obtenus dans tous les cas présentés dans le tableau 3.1. La comparaison des tableaux 2.3 et 3.1 montre des résultats similaires, ce qui indique que la méthode de cryptage a conservé les informations importantes dont le réseau a besoin et que la méthode de classification par génération est capable de traiter les données cryptées. Dans le cas de l'AE, les performances obtenues sur les données claires sont légèrement supérieures à celles obtenues sur les données cryptées.

Les graphiques montrant l'évolution des valeurs de PC en fonction du nombre d'images

TABLE 3.1 – Les précisions de classification (PC %) et les écarts types (±) pour les différents réseaux utilisés (pix2pix et auto-encodeur), pour différentes valeurs de (a, t), sur les données cryptées.

Réseau	**(a, t)**	(5, 5)	(6, 4)	(7, 3)	(8, 2)	(9, 1)
pix2pix	**PC**	94.45	97.31	98.83	99.12	98.75
	Écart-type	2.079	1.282	0.408	0.8	1.677
AE	**PC**	94.5	97	98.08	98.25	98.5
	Écart-type	2.037	1.11	0.917	1.601	2.549

d'apprentissage a par classe pour les deux cas de référence de pix2pix et pour l'auto-encodeur sont présentés dans les figures 3.4a et 3.4b respectivement. Les triangles rouges et verts indiquent respectivement les valeurs maximales et minimales.



(a) pix2pix (b) Auto-encodeur

FIGURE 3.4 – Précision de classification moyenne avec les valeurs maximales et minimales en utilisant les réseaux pix2pix et auto-encodeur pour différentes valeurs de a, sur des données cryptées.

La figure 3.5 montre quelques images générées par les deux générateurs pix2pix et AE. Cette figure montre l'image initiale non cryptée, l'image test cryptée entrée du réseau, la référence ou la sortie souhaitée, et les images générées pour différentes valeurs de (a, t).

La différence de couleur entre les données générées et les données souhaitées est remarquable, mais en général, le réseau a réussi à régénérer des séquences similaires. Cette différence de couleur peut être due à une simple différence de valeurs dans l'un des canaux de couleur. La qualité des données générées s'améliore avec l'augmentation du nombre d'images utilisées pour l'apprentissage.

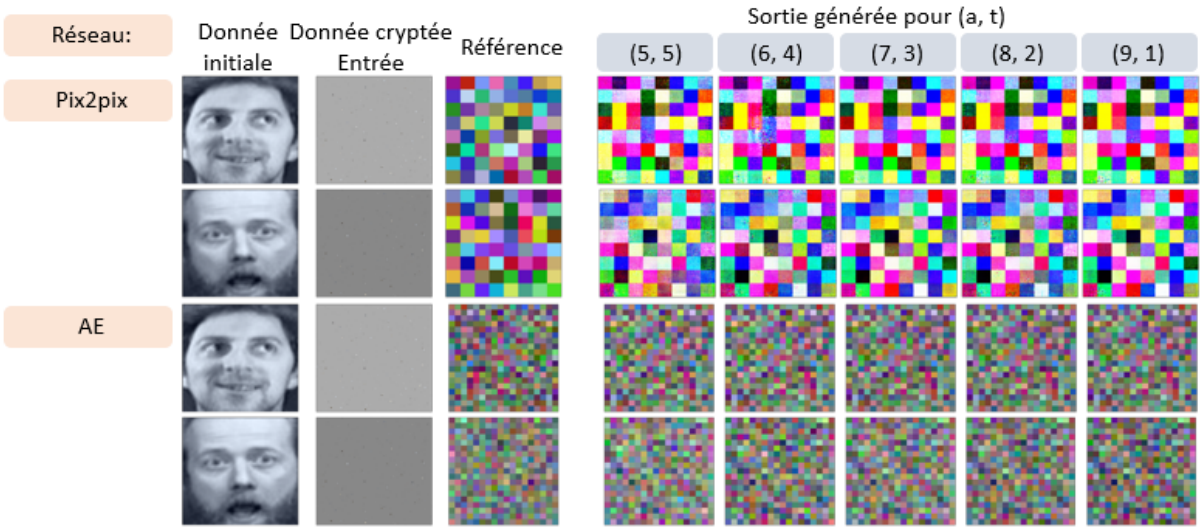


FIGURE 3.5 – Les résultats obtenus pour différentes valeurs (a, t) par les deux générateurs pix2pix et AE ayant comme entrée l'image test cryptée, ainsi que les références

Malgré la grande similitude de toutes les données d'entrée cryptées et l'impossibilité de les distinguer à l'œil nu, le réseau génératif a réussi à générer l'image de référence correcte correspondante à la classe d'entrée. En général, la méthode de cryptage n'a pas vraiment affecté les performances. La méthode de cryptage a conservé toutes les informations nécessaires à la classification des données, sans qu'il soit nécessaire de modifier les réseaux. Cette méthode de cryptage peut donc être considérée comme une méthode perceptuelle. Ces tests ont également montré la capacité de la méthode de classification par génération à traiter différents types de données.

L'application de contrôle d'accès est aussi testée sur les données cryptées. Le tableau 3.2 montre les résultats obtenus pour différents nombres de classes exclues ne, en utilisant les réseaux de génération pix2pix et AE sur des données cryptées. Toutes les valeurs des paramètres mesurant la performance correspondent aux valeurs moyennes de 40 répétitions, en changeant à chaque fois les classes inconnues. Dans tous les cas, les réseaux ont pu classer tous les individus appartenant aux classes connues, indépendamment de la tâche de contrôle d'accès et du seuil à définir.

La figure 3.6 montre quelques exemples d'images générées par pix2pix, et quelques exemples d'images représentatives de la sortie AE, pour différentes images d'entrée cryptées, connues et inconnues, en face des références, dans le cas de $ne = 3$. Dans le cas d'une personne sans images de référence, c'est-à-dire une personne inconnue, le résultat généré

TABLE 3.2 – Les paramètres (PC, PT, PE, PS en %) mesurant les performances des réseaux générateurs, pix2pix et AE, pour différents nombres *ne* de classes exclues de la base d'apprentissage, et pour les seuils choisis, sur des données cryptées.

Réseau	**ne**	1	3	5	10	20
pix2pix	**PC**	100	100	100	100	100
	seuil	0.73	0.734	0.712	0.756	0.752
	PT	95.76	95.68	96.21	94.25	93.25
	PE	95.75	96.08	96.1	94.25	93.16
	PS	95.76	95.86	96.15	94.25	93.17
AE	**PC**	100	100	100	100	100
	seuil	0.487	0.46	0.488	0.516	0.589
	PT	92.2	93.8	93.2	93.3	93.5
	PE	92.25	93.75	93.4	92.97	93.54
	PS	92.24	93.77	93.32	93.04	93.53

peut être considéré comme la combinaison de plusieurs références existantes.

On constate que dans le cas de pix2pix, les résultats obtenus sur les données cryptées sont supérieurs à ceux obtenus sur les données en clair, tandis que dans le cas d'AE, les résultats obtenus sur les données en clair sont supérieurs à ceux obtenus sur les données cryptées. D'où l'impossibilité de déterminer une règle générale.

Cette disparité dans les résultats obtenus peut être considérée comme un point positif de la méthode de classification par génération. Cette méthode peut être utilisée avec n'importe quel réseau et n'importe quelles données. Elle peut donc être considérée comme généraliste, comme les données d'entrée ne l'affectent pas de manière significative.

3.3 Tentative d'amélioration des résultats

La méthode de cryptage dissimule les données originales en y apportant des modifications significatives. Toutes les données d'entrée cryptées sont similaires et la base de données utilisée, ORL, ne contient pas un grand nombre d'images, ni une grande variété de classes. Il est donc intéressant d'entraîner le réseau classificateur avec une grande base de données contenant une grande variété d'images. Une fois l'entraînement terminé, les connaissances acquises peuvent être transférées et adaptées à la petite base de données qui nous intéresse (ORL) en appliquant une méthode d'apprentissage par transfert, le

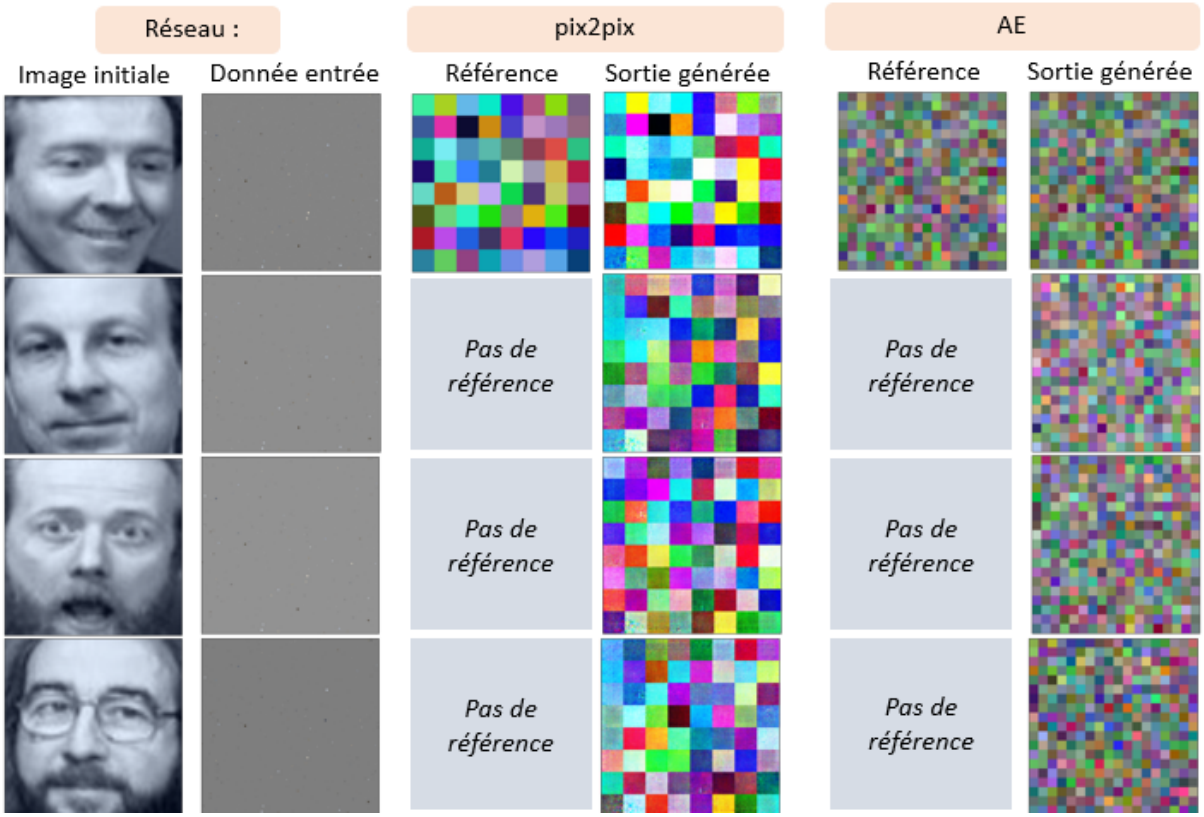


FIGURE 3.6 – Les résultats générés par les deux réseaux générateurs pix2pix et AE pour différentes images d'entrée cryptées connues et inconnues, dans le cas de $ne = 3$.

fine-tuning FT.

La base de données VGG face 2 contenant 1071283 images de 8693 personnes différentes est utilisée comme base de données de pré-entraînement. Toutes les images sont cryptées à l'aide de la nouvelle méthode de cryptage et de la même clé de cryptage K_1.

3.3.1 Entraînement

Dans le cas de pix2pix, les réseaux du générateur et du discriminateur sont entraînés à l'aide de données cryptées de la base de données VGG, en créant une référence aléatoire pour chaque classe, comme décrit dans la section 2.6.3.2. Une fois l'entraînement terminé, le réseau discriminateur est supprimé et le réseau générateur entraîné est ré-entraîné sur les données cryptées de la base de données ORL, après avoir créé la base de données de référence.

Dans le cas de l'auto-encodeur, l'encodeur est entraîné avec le classificateur de produits sphériques [80] sur les données cryptées de la base de données VGG, de sorte que les sorties générées par l'encodeur sont correctement classées par le classificateur. Une fois l'entraînement terminé, le classificateur est supprimé et seul le codeur est ré-entraîné sur les données ORL cryptées, après la création de la base de données de référence.

3.3.2 Création de la base de références

Les connaissances acquises par le réseau lors du pré-entraînement doivent être utilisées pour créer la base de références de la petite base de données d'intérêt. Dans ce cas, le réseau pré entraîné est le responsable de la création de la base de référence, les références ne seront donc plus des références aléatoires, mais des références générées par le réseau en fonction de ce qu'il a appris.

Une image par classe de la base de données ORL est choisie pour être traitée par le réseau générateur pré-entraîné, elle est cryptée en utilisant une clé de cryptage K, et la sortie générée est considérée comme la référence pour sa classe. Dans le cas de pix2pix, l'image générée est constituée de blocs de couleur. Afin de créer correctement la référence, la sortie est régénérée en remplaçant les valeurs de chaque bloc de couleur par la valeur moyenne du bloc (voir figure 3.7). Dans le cas de l'AE, la sortie est un vecteur qui est directement utilise comme référence.

Dans le cas de pix2pix où $K = K_1$, les sorties générées ressemblent au bruit comme le montre la figure 3.8a, ce qui rend impossible la création d'une base de référence, même

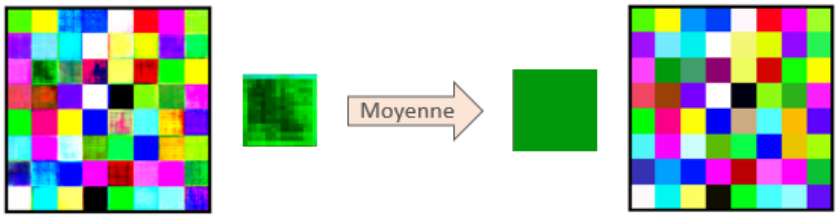
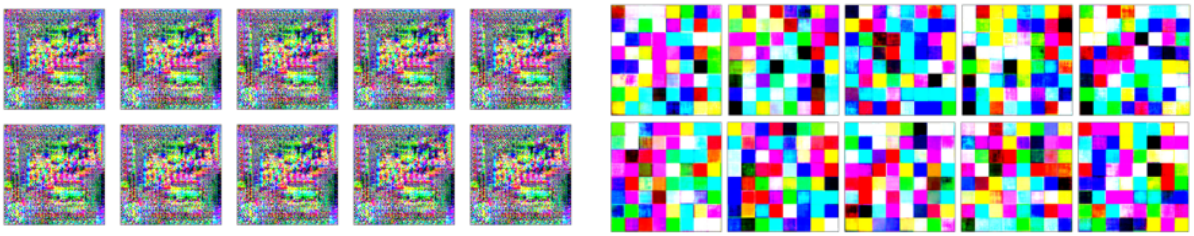


FIGURE 3.7 – Régénération des références en remplaçant les valeurs de chaque bloc de couleur par sa valeur moyenne.

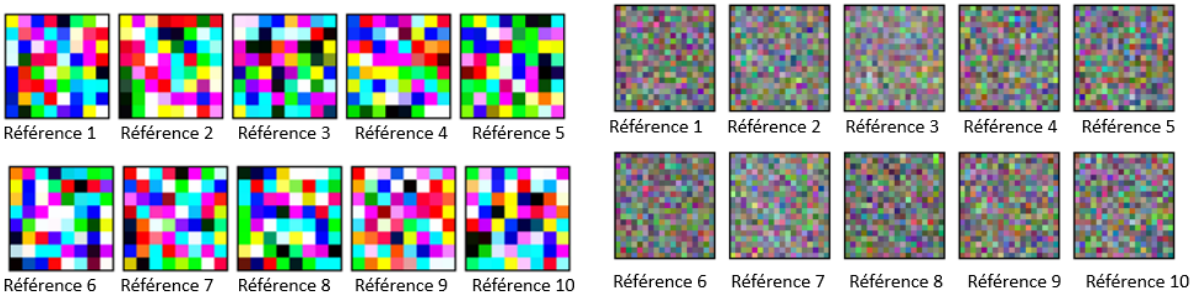
après la régénération des sorties. En changeant la clé de cryptage utilisée pour crypter les images ORL, soit $K = K_2$, le problème est résolu et les images générées sont des blocs de couleurs (voir la figure 3.8b), combinant plusieurs références de la base de données VGG, mais loin de chacune d'entre elles. Avec l'AE, ce problème de création de base de références n'existe pas, car la sortie générée est un vecteur, et non une image. Tous les vecteurs générés sont différents et uniques.



(a) Cryptage par K_1 (b) Cryptage par K_2

FIGURE 3.8 – Les images générées par le réseau pix2pix pré-entraîné sur la base de données VGG cryptée par K_1 de la première classe de la base de données ORL, avant régénération.

Afin de créer la base de références de la base de données ORL, toutes ses images cryptées par $K = K_2$ sont traitées par le réseau pré-entraîné. Pour chaque classe, la sortie générée la plus éloignée (avec le plus grand MSE) de toutes les images générées à partir de l'ensemble de la base ORL est considérée comme la référence de la classe. Les figures 3.9a et 3.9b montrent les références créées par le réseau pré-entraîneur pour les 10 premières classes de la base de données ORL, en utilisant respectivement pix2pix et AE comme générateurs. Dans le cas de pix2pix, les références montrées sont après régénération, et dans le cas d'AE, les références montrées sont les images représentatives des vecteurs.



(a) pix2pix (b) AE

FIGURE 3.9 – Références créées par les réseaux pré-entraînés des dix premières classes de la base de données ORL.

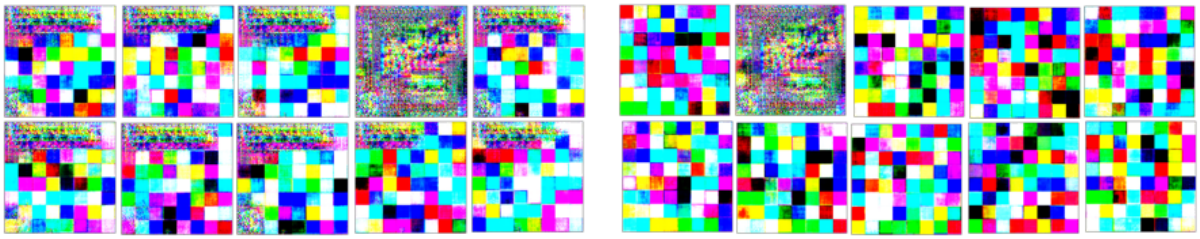
Génération du problème des références

Dans le cas du réseau pix2pix, un problème a été rencontré lors de la création de la base de données de référence. Il a été impossible de créer une base de données de référence en cryptant les images de la base ORL avec la même clé K_1 utilisée pour crypter la base VGG avec laquelle le réseau pix2pix a été pré-entraîné. Les images générées étaient bruitées. Ce problème a été résolu en changeant la clé utilisée pour crypter les images ORL.

Afin de généraliser et d'interpréter ce problème, un deuxième essai a été réalisé avec une autre base de données. La base de données ORL a été remplacée par la base de données ISEN-25 créée dans le cadre de cette thèse. ISEN-25 est une base de données contenant les visages de 25 personnes de l'unité de recherche ISEN. Elle comprend 25 classes, chaque classe étant constituée de 10 images du visage d'une même personne prises sous différents angles. Cette base de données est en couleur.

Le réseau utilisé est pré-entraîné par la base de données VGG cryptée avec la clé K_1. Le réentraînement est effectué à l'aide de la base de données ISEN-25. La première fois, ISEN-25 est cryptée à l'aide de la même clé K_1. Les images générées sont à nouveau du bruit, comme le montre la figure 3.10a, qui rend impossible la création de la base de références. La deuxième fois, ISEN-25 est crypté à l'aide de la deuxième clé de cryptage K_2. Au moins une image générée par classe n'est pas bruitée et peut être considérée comme la référence de la classe. La figure 3.10b montre un exemple de toutes les images générées pour une classe.

D'après les résultats obtenus avec pix2pix, et dans le cas d'un pré-entraînement sur une grande base de données cryptée avec une clé A, le résultat obtenu en utilisant une image d'une autre base de données cryptée avec la même clé A semble bruité. Alors que



(a) Cryptage par K_1 (b) Cryptage par K_2

FIGURE 3.10 – Les images générées par le réseau pix2pix pré-entraîné sur la base de données VGG cryptée par K_1 d'une classe de la base de données ISEN-25, avant régénération.

l'utilisation d'une autre clé de cryptage B donne des résultats de plus en plus proches des résultats souhaités dans la forme, mais très éloignés des références initiales. Ce résultat est en fait un bonus. Dans le cas du "fine tuning", il n'est pas nécessaire d'envoyer la clé A à tous les utilisateurs qui utiliseront le réseau pré-entraîné, mais chaque utilisateur doit créer sa propre clé de cryptage afin de ré-entraîner le réseau sur sa propre base et pour sa propre application.

3.3.3 Résultats

Le réseau ré-entraîné avec la base de données ORL cryptée utilisant la clé K_2 est testé dans l'application de contrôle d'accès. Les réseaux pix2pix et AE sont utilisés. Le tableau 3.3 montre les résultats obtenus pour différents nombres de classes exclues ne. Toutes les valeurs des paramètres mesurant la performance correspondent aux valeurs moyennes de 40 répétitions, en changeant à chaque fois les classes inconnues.

En comparant les tableaux 3.2 et 3.3, il apparaît que dans le cas de pix2pix, les résultats sans FT sont meilleurs que ceux avec FT, et que dans le cas d'AE, les résultats avec FT sont meilleurs que ceux sans FT. Ces résultats peuvent être dus à la nature différente de la sortie des deux générateurs. Avec une sortie vectorielle, la génération devient plus facile qu'avec une sortie image, et avec un changement de base de données, l'adaptation à la nouvelle base de données et à la nouvelle application devient plus facile. Dans le cas de l'AE, il n'y a pas de grand changement de référence. Les références dans ce cas sont des vecteurs dont les valeurs changeront, alors que dans le cas de pix2pix les références sont des images ; l'image entière, les bordures, les couleurs et la séquence des couleurs doivent être adaptées.

TABLE 3.3 – Les paramètres (PC, PT, PE, PS en %) mesurant les performances des réseaux générateurs, pix2pix et AE, pour différents nombres *ne* de classes exclues de la base d'apprentissage, et pour les seuils choisis, sur des données cryptées après ré-entraînement.

Réseau	**ne**	1	3	5	10	20
pix2pix	**PC**	99.94	100	99.71	100	99.75
	seuil	0.415	0.458	0.366	0.479	0.3
	PT	95.64	96.62	96.86	96.75	94.75
	PE	95.5	96.58	96.6	96.77	94.7
	PS	95.61	96.6	96.1	96.77	94.7
AE	**PC**	100	100	100	100	100
	seuil	0.3	0.306	0.307	0.322	0.355
	PT	95.5	96.4	95.1	94.8	95.1
	PE	95.5	96.67	95.05	94.5	95.19
	PS	95.46	96.53	95.09	94.71	95.18

3.4 Test de robustesse

Avec le développement intensif des méthodes d'attaque, il est essentiel que le système développé soit soumis à des tests de robustesse pour vérifier son efficacité et garantir la sécurité des données privées sensibles. Le système de classification par génération des données cryptées est soumis à des tests de robustesse afin de déterminer la fiabilité de ses résultats et de tester la robustesse de la méthode de cryptage face à la modification des clés de cryptage.

Pour tester la robustesse du système, le réseau générateur est entraîné sur la base de données ORL cryptée à l'aide de la clé K_1, et testé sans ré-entraînement sur une autre base de données BT. L'idéal est que le système classifie toute la base de données BT comme inconnue (ATN = 100%). Le tableau 3.4 montre les résultats obtenus pour différents cas de BT, et en utilisant les deux réseaux générateurs, le pix2pix et l'AE. Ces résultats sont obtenus en fixant le seuil de décision à 0.5.

Le classifieur a pu rejeter toutes les bases de données ORL et ISEN-5 non cryptées, ce qui indique que la méthode de cryptage a effectivement changé et caché les informations importantes. De même, en changeant la clé de cryptage, le classifieur n'a pas pu reconnaître les mêmes personnes ou de nouvelles personnes, et il les a classées comme inconnues, malgré la similitude générale de forme et de couleur entre les données cryptées

TABLE 3.4 – Résultats des tests de robustesse sur la base BT, avec un seuil égal à 0.5.

BT	Réseau	ATN
ORL non cryptée	pix2pix	100
	AE	100
ORL cryptée K_2	pix2pix	100
	AE	100
ISEN-25 non cryptée	pix2pix	100
	AE	100
ISEN-25 cryptée K_1	pix2pix	91.6
	AE	100
ISEN-25 cryptée K_2	pix2pix	100
	AE	100

par deux clés différentes. Ce résultat montre l'importance de la clé de cryptage et la difficulté de récupérer les informations importantes sans elle. Dans le cas de l'utilisation de la même clé de cryptage sur une autre base de données, les réseaux ont classé la plupart des images comme inconnues, mais le pix2pix a fait ressembler certaines images à des personnes connues.

3.5 Conclusion

Dans ce chapitre, une nouvelle méthode de cryptage basée sur la DCT est présentée. Cette méthode de cryptage transforme les données claires dans le domaine fréquentiel en appliquant la DCT, puis, à l'aide de la clé de cryptage privée composée de deux éléments, des modifications sont apportées en appliquant des opérations de calcul et en réarrangeant et en mélangeant les données. Cette méthode de cryptage modifie le domaine et les données importantes, ce qui donne une image grise avec peu d'informations, qui semble difficile à traiter. L'avantage de cette méthode de cryptage est la possibilité de la réaliser optiquement et la facilité de l'utiliser dans le système de classification développé dans le chapitre précédent.

La méthode de cryptage présentée dans ce chapitre peut être considérée comme une méthode de cryptage perceptuelle, car elle ne nécessite aucune modification du réseau de neurones utilisé pour traiter les données cryptées. Cette méthode de cryptage a été utilisée pour dissimuler des informations privées des données utilisées dans la méthode

de classification par génération. Elle n'a pas affecté les performances de classification et de contrôle d'accès, et le classificateur a été capable de correctement classer les données cryptées et de distinguer les personnes inconnues dans le domaine crypté. La méthode de classification par génération et la méthode de cryptage forment ensemble un système d'intelligence artificielle qui préserve la confidentialité des données privées sensibles.

Une tentative d'amélioration des performances a été testée en utilisant le "Fine Tuning" (réglage fin). Cette méthode a permis d'améliorer les performances de l'AE mais n'a pas donné les résultats attendus avec le pix2pix, probablement en raison de la différence de nature de la sortie générée par les deux générateurs. Le FT a été utilisé pour transférer les connaissances acquises pendant l'entraînement sur une grande base de données cryptée, en créant des références basées sur les connaissances acquises et en utilisant le modèle pré-entraîné comme point de départ pour le ré-entraînement sur la base de données à utiliser.

La robustesse de ce système a été évaluée en le testant sur des bases de données différentes de celle utilisée pour l'entraînement. Ce test de robustesse a montré l'efficacité de la méthode de cryptage et l'importance de la clé de cryptage.

Conclusion générale et perspectives

Cette thèse cherche à résoudre l'un des principaux problèmes auxquels est confrontée l'application de l'intelligence artificielle dans certains domaines sensibles. Son objectif est de développer un système d'intelligence artificielle qui préserve la confidentialité des données privées utilisées. Cette préservation est obtenue en classant les données cryptées dans des applications de reconnaissance faciale et de contrôle d'accès.

Le système développé dans cette thèse se compose de deux parties principales : la classification et le cryptage. La combinaison de ces deux parties indépendantes a permis la réalisation d'un système de classification sécurisé de bout en bout.

Une nouvelle méthode de classification par génération a été développée et adaptée, qui utilise un réseau de neurones capable de générer des données en sortie, appelé dans ce manuscrit réseau générateur. Tout réseau présentant cette caractéristique peut être utilisé, mais dans ce travail nous avons appliqué le générateur de pix2pix et l'encodeur de l'auto-encodeur.

Cette méthode nécessite de plus la création d'une base de données de référence préliminaire, contenant des données présentant toutes les classes d'entraînement. Différents choix de base de référence ont été testés, et il a été constaté que le choix de la référence présentant la classe n'affectait pas les performances du classificateur. Ainsi, les références peuvent être choisies soit dans la base d'entraînement si cela est possible (dans le cas où la sortie du générateur est de même nature que son entrée), soit de manière aléatoire.

Enfin, pour que cette méthode de classification prenne la décision de la classe appropriée pour l'image d'entrée, une comparaison est effectuée entre l'image de sortie générée par le générateur et toutes les images de référence. Diverses applications de la reconnaissance de chiffres manuscrits de MNIST, de la reconnaissance faciale et du contrôle d'accès de l'ORL ont été testées, et ont donné de bons résultats. Il convient de noter que la base de données ORL a été choisie en raison de sa petite taille, ce qui la rend adaptée à ce type de travail, qui vise à contrôler l'accès à la gestion d'une entreprise ou d'une école, par exemple, qui compte un petit nombre de personnes. Le personnel administratif ne devrait

pas dépasser quelques dizaines de personnes. D'où le choix de cette base de données. La méthode de classification développée peut être utilisée avec des données en clair, ou avec des données cryptées, pour des applications qui préservent la sécurité de ces données.

La deuxième partie du système global développé dans cette thèse est la méthode de cryptage, qui garantira la préservation de la confidentialité des données personnelles privées. Cette méthode de cryptage est une méthode optique qui peut être appliquée directement dans la caméra lors de la saisie de l'image, basée sur la DCT. Deux clés de cryptage sont nécessaires, l'une utilisée avec la DCT de l'image pour obtenir un spectre crypté, et l'autre utilisée pour réarranger les données cryptées afin de cacher des informations telles que l'utilisation de la DCT et la taille du bloc de la DCT. Les données cryptées à l'aide de cette méthode de cryptage peuvent être utilisées directement auprès du classificateur déjà développé, sans qu'il soit nécessaire de modifier ce dernier. L'utilisation des données cryptées n'affecte pas considérablement les performances du classificateur et du contrôleur d'accès ; au contraire, dans certains cas, de meilleurs résultats sont obtenus.

Une tentative pour améliorer les résultats a été faite par apprentissage de transfert avec la grande base de données de VGG Face 2, mais l'amélioration souhaitée de la performance n'a pas été atteinte. D'autre part, ce système a démontré un haut degré de robustesse en changeant la clé de cryptage et la base de données d'entraînement, ce qui met en évidence l'efficacité de la méthode de cryptage et la validité de la clé de cryptage. D'autres tests de robustesse pourront être effectués.

Cette thèse a abouti à des publications, des présentations à des conférences, des séminaires et d'autres journées de recherche :

— Un article de journal sur l'état de l'art qui traite des méthodes de préservation de la sécurité des données utilisées par les réseaux de neurones.
R. El Saj, E. Sedgh Gooya, A. Alfalou et M. Khalil, « Privacy-preserving deep neural network methods : computational and perceptual methods—an overview », Electronics, t. 10, 11, p. 1367, 2021 [38].

— Un article de conférence présentant la méthode de classification par génération.
R. El Saj, E. S. Gooya, A. Alfalou et M. Khalil, « Generative classifier pix2pix », in Pattern Recognition and Tracking XXXIV, SPIE, t. 12527, 2023, p. 85-92 [108]

— Un article de conférence sur l'application de la reconnaissance faciale et du contrôle d'accès par la méthode de classification par génération.
R. El Saj, E. S. Gooya, A. Alfalou et M. Khalil, « Face recognition and access control applications of the generative classifier pix2pix », in Pattern Recognition

and Tracking XXXIV, SPIE, t. 12527, 2023, p. 60-66 [123].

— un article de conférence, en vue de sa publication, présentant la méthode de cryptage et la classification par génération utilisant l'encodeur sur des données cryptées.
R. El Saj, C. H. Pham, E. S. Gooya, A. Alfalou et M. Khalil, « Privacy Preserving Encoder Classifier for Access Control Based on Face Recognition », in 2023 Twelfth International Conference on Image Processing Theory, Tools and Applications (IPTA).

Les travaux développés dans cette thèse ouvrent des perspectives et des possibilités de recherche supplémentaires. L'une des perspectives de ce système est la possibilité d'ajouter une nouvelle classe sans nécessiter de ré-entraînement. Étant donné que le générateur est chargé de créer des sorties spécifiques à une classe, l'objectif est que, pour toute nouvelle classe à ajouter, le générateur n'ait besoin que d'une seule image de la nouvelle classe. La classification doit être inconnue. La sortie est alors ajoutée à la base de référence et considérée comme la référence de cette classe. Si une deuxième image de cette nouvelle classe est traitée par le réseau, la sortie générée doit être très similaire à la référence déjà ajoutée, la décision doit être connue et l'image doit être classée comme appartenant à la nouvelle classe ajoutée.

En outre, l'efficacité et la robustesse de la méthode de cryptage peuvent être vérifiées en effectuant des tests spécifiques. De même, la robustesse du système global peut être vérifiée en effectuant des tests supplémentaires contre les attaques de l'adversaire ou les attaques du réseau de transformation inverse.

BIBLIOGRAPHIE

[1] J. McCarthy, M. L. Minsky, N. Rochester et C. E. Shannon, « A proposal for the dartmouth summer research project on artificial intelligence, august 31, 1955 », *AI magazine*, t. 27, *4*, p. 12-12, 2006.

[2] J. Weizenbaum, « ELIZA—a Computer Program for the Study of Natural Language Communication between Man and Machine », *Commun. ACM*, t. 9, *1*, p. 36-45, jan. 1966, issn : 0001-0782. doi : 10.1145/365153.365168. adresse : https://doi.org/10.1145/365153.365168.

[3] A. Newell, J. C. Shaw et H. A. Simon, « Report on a general problem solving program », in *IFIP congress*, Pittsburgh, PA, t. 256, 1959, p. 64.

[4] M. Haenlein et A. Kaplan, « A Brief History of Artificial Intelligence : On the Past, Present, and Future of Artificial Intelligence », *California Management Review*, t. 61, *4*, p. 5-14, 2019. doi : 10.1177/0008125619864925. eprint : https://doi.org/10.1177/0008125619864925. adresse : https://doi.org/10.1177/0008125619864925.

[5] K. Fukushima, « Neocognitron : A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position », *Biological cybernetics*, t. 36, *4*, p. 193-202, 1980.

[6] D. E. Rumelhart, G. E. Hinton et R. J. Williams, « Learning representations by back-propagating errors », *nature*, t. 323, *6088*, p. 533-536, 1986.

[7] A. L. Fradkov, « Early history of machine learning », *IFAC-PapersOnLine*, t. 53, *2*, p. 1385-1390, 2020.

[8] P. P. Shinde et S. Shah, « A review of machine learning and deep learning applications », in *2018 Fourth international conference on computing communication control and automation (ICCUBEA)*, IEEE, 2018, p. 1-6.

[9] Z.-Q. Zhao, P. Zheng, S.-t. Xu et X. Wu, « Object detection with deep learning : A review », *IEEE transactions on neural networks and learning systems*, t. 30, *11*, p. 3212-3232, 2019.

[10] Y. Xiao, Z. Tian, J. Yu et al., « A review of object detection based on deep learning », *Multimedia Tools and Applications*, t. 79, p. 23 729-23 791, 2020.

[11] I. Kononenko, I. Bratko et M. Kukar, « Application of machine learning to medical diagnosis », *Machine learning and data mining : Methods and applications*, t. 389, p. 408, 1997.

[12] M. Y. Shaheen, « Adoption of machine learning for medical diagnosis », *ScienceOpen Preprints*, 2021.

[13] N. H. Tandel, H. B. Prajapati et V. K. Dabhi, « Voice recognition and voice comparison using machine learning techniques : A survey », in *2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS)*, IEEE, 2020, p. 459-465.

[14] M. Buyukyilmaz et A. O. Cibikdiken, « Voice gender recognition using deep learning », in *2016 International Conference on Modeling, Simulation and Optimization Technologies and Applications (MSOTA2016)*, Atlantis Press, 2016, p. 409-411.

[15] B. Balci, D. Saadati et D. Shiferaw, « Handwritten text recognition using deep learning », *CS231n : Convolutional Neural Networks for Visual Recognition, Stanford University, Course Project Report, Spring*, p. 752-759, 2017.

[16] M. T. H. Fuad, A. A. Fime, D. Sikder et al., « Recent advances in deep learning techniques for face recognition », *IEEE Access*, t. 9, p. 99 112-99 142, 2021.

[17] H. Ge, Z. Zhu, Y. Dai, B. Wang et X. Wu, « Facial expression recognition based on deep learning », *Computer Methods and Programs in Biomedicine*, t. 215, p. 106 621, 2022.

[18] A. M. Ozbayoglu, M. U. Gudelek et O. B. Sezer, « Deep learning for financial applications : A survey », *Applied Soft Computing*, t. 93, p. 106 384, 2020.

[19] K. Y. Chan, B. Abu-Salih, R. Qaddoura et al., « Deep Neural Networks in the Cloud : Review, Applications, Challenges and Research Directions », *Neurocomputing*, p. 126 327, 2023.

[20] M. Zviran et Z. Erlich, « Identification and authentication : technology and implementation issues », *Communications of the Association for Information Systems*, t. 17, *1*, p. 4, 2006.

[21] Y. Spector et J. Ginzberg, « Pass-sentence—a new approach to computer code », *Computers & Security*, t. 13, *2*, p. 145-160, 1994.

[22] S. Wiedenbeck, J. Waters, J.-C. Birget, A. Brodskiy et N. Memon, « PassPoints : Design and longitudinal evaluation of a graphical password system », *International journal of human-computer studies*, t. 63, *1-2*, p. 102-127, 2005.

[23] D. Davis, F. Monrose et M. K. Reiter, « On user choice in graphical password schemes. », in *USENIX security symposium*, t. 13, 2004, p. 11-11.

[24] M. Zviran et W. J. Haga, « A comparison of password techniques for multilevel authentication mechanisms », *The Computer Journal*, t. 36, *3*, p. 227-237, 1993.

[25] S. Z. S. Idrus, E. Cherrier, C. Rosenberger et J.-J. Schwartzmann, « A review on authentication methods », *Australian Journal of Basic and Applied Sciences*, t. 7, *5*, p. 95-107, 2013.

[26] Z. Rui et Z. Yan, « A survey on biometric authentication : Toward secure and privacy-preserving identification », *IEEE access*, t. 7, p. 5994-6009, 2018.

[27] D. Bhattacharyya, R. Ranjan, F. Alisherov, M. Choi et al., « Biometric authentication : A review », *International Journal of u-and e-Service, Science and Technology*, t. 2, *3*, p. 13-28, 2009.

[28] N. Singla, M. Kaur et S. Sofat, « Automated latent fingerprint identification system : A review », *Forensic science international*, t. 309, p. 110 187, 2020.

[29] J. Hájek et M. Drahanskỳ, « Recognition-based on eye biometrics : Iris and retina », *Biometric-Based Physical and Cybersecurity Systems*, p. 37-102, 2019.

[30] N. D. AL-Shakarchy, H. K. Obayes et Z. N. Abdullah, « Person identification based on voice biometric using deep neural network », *International Journal of Information Technology*, p. 1-7, 2022.

[31] Y. Kortli, M. Jridi, A. Al Falou et M. Atri, « Face recognition systems : A survey », *Sensors*, t. 20, *2*, p. 342, 2020.

[32] M. Zulfiqar, F. Syed, M. J. Khan et K. Khurshid, « Deep face recognition for biometric authentication », in *2019 international conference on electrical, communication, and computer engineering (ICECCE)*, IEEE, 2019, p. 1-6.

[33] H. Hanizan, R. Din, A. Hafiza, A. Rukhiya et M. Noor, « A review of artificial intelligence techniques in image steganography domain », *Journal of Engineering Science and Technology*, p. 106-116, 2017.

[34] J. Liu, Y. Ke, Z. Zhang et al., « Recent advances of image steganography with generative adversarial networks », *IEEE Access*, t. 8, p. 60 575-60 597, 2020.

[35] M. Chaumont, « Deep learning in steganography and steganalysis », in *Digital media steganography*, Elsevier, 2020, p. 321-349.

[36] Y. Xiao, Q. Hao et D. D. Yao, « Neural cryptanalysis : metrics, methodology, and applications in CPS ciphers », in *2019 IEEE conference on dependable and secure computing (DSC)*, IEEE, 2019, p. 1-8.

[37] E. Volna, M. Kotyrba, V. Kocian et M. Janosek, « Cryptography based on neural network. », in *ECMS*, 2012, p. 386-391.

[38] R. El Saj, E. Sedgh Gooya, A. Alfalou et M. Khalil, « Privacy-preserving deep neural network methods : computational and perceptual methods—an overview », *Electronics*, t. 10, *11*, p. 1367, 2021.

[39] H. C. Tanuwidjaja, R. Choi, S. Baek et K. Kim, « Privacy-preserving deep learning on machine learning as a service—a comprehensive survey », *IEEE Access*, t. 8, p. 167 425-167 447, 2020.

[40] A. Acar, H. Aksu, A. S. Uluagac et M. Conti, « A survey on homomorphic encryption schemes : Theory and implementation », *ACM Computing Surveys (Csur)*, t. 51, *4*, p. 1-35, 2018.

[41] D. Boneh, A. Sahai et B. Waters, « Functional encryption : Definitions and challenges », in *Theory of Cryptography : 8th Theory of Cryptography Conference, TCC 2011, Providence, RI, USA, March 28-30, 2011. Proceedings 8*, Springer, 2011, p. 253-273.

[42] R. Gilad-Bachrach, N. Dowlin, K. Laine, K. Lauter, M. Naehrig et J. Wernsing, « Cryptonets : Applying neural networks to encrypted data with high throughput and accuracy », in *International conference on machine learning*, PMLR, 2016, p. 201-210.

[43] E. Hesamifard, H. Takabi et M. Ghasemi, « Cryptodl : Deep neural networks over encrypted data », *arXiv preprint arXiv :1711.05189*, 2017.

[44] R. Xu, J. B. Joshi et C. Li, « Cryptonn : Training neural networks over encrypted data », in *2019 IEEE 39th International Conference on Distributed Computing Systems (ICDCS)*, IEEE, 2019, p. 1199-1209.

[45] T. Maekawa, A. Kawamura, Y. Kinoshita et H. Kiya, « Privacy-preserving svm computing in the encrypted domain », in *2018 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, IEEE, 2018, p. 897-902.

[46] A. Kawamura, Y. Kinoshita, T. Nakachi, S. Shiota et H. Kiya, « A privacy-preserving machine learning scheme using etc images », *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, t. 103, *12*, p. 1571-1578, 2020.

[47] M. Tanaka, « Learnable Image Encryption », in *2018 IEEE International Conference on Consumer Electronics-Taiwan (ICCE-TW)*, 2018, p. 1-2. doi : 10.1109/ICCE-China.2018.8448772.

[48] W. Sirichotedumrong et H. Kiya, « Visual security evaluation of learnable image encryption methods against ciphertext-only attacks », in *2020 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, IEEE, 2020, p. 1304-1309.

[49] W. Sirichotedumrong, T. Maekawa, Y. Kinoshita et H. Kiya, « Privacy-preserving deep neural networks with pixel-based image encryption considering data augmentation in the encrypted domain », in *2019 IEEE International Conference on Image Processing (ICIP)*, IEEE, 2019, p. 674-678.

[50] W. Sirichotedumrong, Y. Kinoshita et H. Kiya, « Pixel-based image encryption without key management for privacy-preserving deep neural networks », *Ieee Access*, t. 7, p. 177 844-177 855, 2019.

[51] W. Sirichotedumrong et H. Kiya, « A gan-based image transformation scheme for privacy-preserving deep neural networks », in *2020 28th European Signal Processing Conference (EUSIPCO)*, IEEE, 2021, p. 745-749.

[52] H. Ito, Y. Kinoshita et H. Kiya, « A framework for transformation network training in coordination with semi-trusted cloud provider for privacy-preserving deep neural networks », in *2020 Asia-Pacific Signal and Information Processing*

Association Annual Summit and Conference (APSIPA ASC), IEEE, 2020, p. 1420-1424.

[53] H. Ito, Y. Kinoshita et H. Kiya, « Image transformation network for privacy-preserving deep neural networks and its security evaluation », in *2020 IEEE 9th Global Conference on Consumer Electronics (GCCE)*, IEEE, 2020, p. 822-825.

[54] I. Adjabi, A. Ouahabi, A. Benzaoui et A. Taleb-Ahmed, « Past, present, and future of face recognition : A review », *Electronics*, t. 9, *8*, p. 1188, 2020.

[55] A. Y. J. Nakanishi et B. J. Western, « Advancing the state-of-the-art in transportation security identification and verification technologies : Biometric and multi-biometric systems », in *2007 IEEE Intelligent Transportation Systems Conference*, IEEE, 2007, p. 1004-1009.

[56] M. Chihaoui, A. Elkefi, W. Bellil et C. Ben Amar, « A survey of 2D face recognition techniques », *Computers*, t. 5, *4*, p. 21, 2016.

[57] M. Turk et A. Pentland, « Eigenfaces for recognition », *Journal of cognitive neuroscience*, t. 3, *1*, p. 71-86, 1991.

[58] M. A. Turk et A. P. Pentland, « Face recognition using eigenfaces », in *Proceedings. 1991 IEEE computer society conference on computer vision and pattern recognition*, IEEE Computer Society, 1991, p. 586-587.

[59] J. H. Shah, M. Sharif, M. Raza et A. Azeem, « A Survey : Linear and Nonlinear PCA Based Face Recognition Techniques. », *Int. Arab J. Inf. Technol.*, t. 10, *6*, p. 536-545, 2013.

[60] K. Simonyan, O. M. Parkhi, A. Vedaldi et A. Zisserman, « Fisher vector faces in the wild. », in *BMVC*, t. 2, 2013, p. 4.

[61] P. N. Belhumeur, J. P. Hespanha et D. J. Kriegman, « Eigenfaces vs. fisherfaces : Recognition using class specific linear projection », *IEEE Transactions on pattern analysis and machine intelligence*, t. 19, *7*, p. 711-720, 1997.

[62] Z. Sufyanu, F. S. Mohamad, A. A. Yusuf et M. B. Mamat, « Enhanced Face Recognition Using Discrete Cosine Transform. », *Engineering Letters*, t. 24, *1*, 2016.

[63] Z.-H. Huang, W.-J. Li, J. Shang, J. Wang et T. Zhang, « Non-uniform patch based face recognition via 2D-DWT », *Image and Vision Computing*, t. 37, p. 12-19, 2015.

[64] L. Wiskott, « Phantom faces for face analysis », *Pattern Recognition*, t. 30, *6*, p. 837-846, 1997.

[65] M. Lades, J. C. Vorbruggen, J. Buhmann et al., « Distortion invariant object recognition in the dynamic link architecture », *IEEE Transactions on computers*, t. 42, *3*, p. 300-311, 1993.

[66] C. Kotropoulos, A. Tefas et I. Pitas, « Frontal face authentication using morphological elastic graph matching », *IEEE Transactions on Image Processing*, t. 9, *4*, p. 555-560, 2000.

[67] Y. Ouerhani, A. Alfalou et C. Brosseau, « Road mark recognition using HOG-SVM and correlation », in *Optics and Photonics for Information Processing XI*, SPIE, t. 10395, 2017, p. 119-126.

[68] L. Shen, L. Bai et Z. Ji, « A svm face recognition method based on optimized gabor features », in *Advances in Visual Information Systems : 9th International Conference, VISUAL 2007 Shanghai, China, June 28-29, 2007 Revised Selected Papers 9*, Springer, 2007, p. 165-174.

[69] A. A. Fathima, S. Ajitha, V. Vaidehi, M. Hemalatha, R. Karthigaiveni et R. Kumar, « Hybrid approach for face recognition combining gabor wavelet and linear discriminant analysis », in *2015 IEEE international conference on computer graphics, vision and information security (CGVIS)*, IEEE, 2015, p. 220-225.

[70] H. Supreetha Gowda, G. Hemantha Kumar et M. Imran, « Multimodal biometric recognition system based on nonparametric classifiers », in *Data Analytics and Learning : Proceedings of DAL 2018*, Springer, 2019, p. 269-278.

[71] D. Pratima et N. Nimmakanti, « Pattern recognition algorithms for cluster identification problem », *Int. J. Comput. Sci. Inform*, t. 1, p. 2231-5292, 2012.

[72] C. Zhao, X. Li et Y. Cang, « Bisecting k-means clustering based face recognition using block-based bag of words model », *Optik-International Journal for Light and Electron Optics*, t. 126, *19*, p. 1761-1766, 2015.

[73] Y. Sun, D. Liang, X. Wang et X. Tang, « Deepid3 : Face recognition with very deep neural networks », *arXiv preprint arXiv :1502.00873*, 2015.

[74] F. Schroff, D. Kalenichenko et J. Philbin, « Facenet : A unified embedding for face recognition and clustering », in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, p. 815-823.

[75] X. Wu, R. He, Z. Sun et T. Tan, « A light CNN for deep face representation with noisy labels », *IEEE Transactions on Information Forensics and Security*, t. 13, *11*, p. 2884-2896, 2018.

[76] X. Zhang, Z. Fang, Y. Wen, Z. Li et Y. Qiao, « Range loss for deep face recognition with long-tailed training data », in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, p. 5409-5418.

[77] H. Ben Fredj, S. Bouguezzi et C. Souani, « Face recognition in unconstrained environment with CNN », *The Visual Computer*, t. 37, p. 217-226, 2021.

[78] F. Wang, X. Xiang, J. Cheng et A. L. Yuille, « Normface : L2 hypersphere embedding for face verification », in *Proceedings of the 25th ACM international conference on Multimedia*, 2017, p. 1041-1049.

[79] B. Chen, W. Deng et J. Du, « Noisy softmax : Improving the generalization ability of dcnn via postponing the early softmax saturation », in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, p. 5372-5381.

[80] J. Deng, J. Guo, N. Xue et S. Zafeiriou, « Arcface : Additive angular margin loss for deep face recognition », in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, p. 4690-4699.

[81] S. B. Kotsiantis, I. Zaharakis, P. Pintelas et al., « Supervised machine learning : A review of classification techniques », *Emerging artificial intelligence applications in computer engineering*, t. 160, *1*, p. 3-24, 2007.

[82] M. A. Chandra et S. Bedi, « Survey on SVM and their application in image classification », *International Journal of Information Technology*, t. 13, p. 1-11, 2021.

[83] G. Guo, H. Wang, D. Bell, Y. Bi et K. Greer, « KNN model-based approach in classification », in *On The Move to Meaningful Internet Systems 2003 : CoopIS, DOA, and ODBASE : OTM Confederated International Conferences, CoopIS, DOA, and ODBASE 2003, Catania, Sicily, Italy, November 3-7, 2003. Proceedings*, Springer, 2003, p. 986-996.

[84] A. A. M. Al-Saffar, H. Tao et M. A. Talab, « Review of deep convolution neural network in image classification », in *2017 International conference on radar, antenna, microwave, electronics, and telecommunications (ICRAMET)*, IEEE, 2017, p. 26-31.

[85] I. Goodfellow, J. Pouget-Abadie, M. Mirza et al., « Generative adversarial nets », *Advances in neural information processing systems*, t. 27, 2014.

[86] D. P. Kingma et M. Welling, « Auto-encoding variational bayes », *arXiv preprint arXiv :1312.6114*, 2013.

[87] C. Ledig, L. Theis, F. Huszár et al., « Photo-realistic single image super-resolution using a generative adversarial network », in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, p. 4681-4690.

[88] Z. Ding, X.-Y. Liu, M. Yin et L. Kong, « Tgan : Deep tensor generative adversarial nets for large image generation », *arXiv preprint arXiv :1901.09953*, 2019.

[89] P. Isola, J.-Y. Zhu, T. Zhou et A. A. Efros, « Image-to-image translation with conditional adversarial networks », in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, p. 1125-1134.

[90] J.-Y. Zhu, T. Park, P. Isola et A. A. Efros, « Unpaired image-to-image translation using cycle-consistent adversarial networks », in *Proceedings of the IEEE international conference on computer vision*, 2017, p. 2223-2232.

[91] C. Wang, C. Xu, C. Wang et D. Tao, « Perceptual adversarial networks for image-to-image transformation », *IEEE Transactions on Image Processing*, t. 27, *8*, p. 4066-4079, 2018.

[92] Y. Pang, J. Lin, T. Qin et Z. Chen, « Image-to-image translation : Methods and applications », *IEEE Transactions on Multimedia*, t. 24, p. 3859-3881, 2021.

[93] S. Frolov, T. Hinz, F. Raue, J. Hees et A. Dengel, « Adversarial text-to-image synthesis : A review », *Neural Networks*, t. 144, p. 187-209, 2021.

[94] T. Xu, P. Zhang, Q. Huang et al., « Attngan : Fine-grained text to image generation with attentional generative adversarial networks », in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, p. 1316-1324.

[95] V. Sampath, I. Maurtua, J. J. Aguilar Martin et A. Gutierrez, « A survey on generative adversarial networks for imbalance problems in computer vision tasks », *Journal of big Data*, t. 8, p. 1-59, 2021.

[96] M. Mirza et S. Osindero, « Conditional generative adversarial nets », *arXiv preprint arXiv :1411.1784*, 2014.

[97] A. Odena, C. Olah et J. Shlens, « Conditional image synthesis with auxiliary classifier gans », in *International conference on machine learning*, PMLR, 2017, p. 2642-2651.

[98] S. Bazrafkan et P. Corcoran, « Versatile auxiliary classifier with generative adversarial network (vac+ gan), multi class scenarios », *arXiv preprint arXiv :1806.07751*, 2018.

[99] D. E. Rumelhart, G. E. Hinton, R. J. Williams et al., *Learning internal representations by error propagation*, 1985.

[100] D. Bank, N. Koenigstein et R. Giryes, « Autoencoders », *arXiv preprint arXiv :2003.05991*, 2020.

[101] M. Sewak, S. K. Sahay et H. Rathore, « An overview of deep learning architecture of deep neural networks and autoencoders », *Journal of Computational and Theoretical Nanoscience*, t. 17, *1*, p. 182-188, 2020.

[102] P. Vincent, H. Larochelle, Y. Bengio et P.-A. Manzagol, « Extracting and composing robust features with denoising autoencoders », in *Proceedings of the 25th international conference on Machine learning*, 2008, p. 1096-1103.

[103] A. Ng et al., « Sparse autoencoder », *CS294A Lecture notes*, t. 72, *2011*, p. 1-19, 2011.

[104] S. Rifai, P. Vincent, X. Muller, X. Glorot et Y. Bengio, « Contractive auto-encoders : Explicit invariance during feature extraction », in *Proceedings of the 28th international conference on international conference on machine learning*, 2011, p. 833-840.

[105] M. Gogoi et S. A. Begum, « Image classification using deep autoencoders », in *2017 IEEE international conference on computational intelligence and computing research (ICCIC)*, IEEE, 2017, p. 1-5.

[106] H. W. L. Mak, R. Han et H. H. Yin, « Application of variational autoEncoder (VAE) model and image processing approaches in game design », *Sensors*, t. 23, *7*, p. 3457, 2023.

[107] E. Sedgh-Gooya et A. Alfalou, « Few-Shot Learning using Supervised Non-Associative Autoencoders and Correlation Techniques »,

[108] R. El Saj, E. S. Gooya, A. Alfalou et M. Khalil, « Generative classifier pix2pix », in *Pattern Recognition and Tracking XXXIV*, SPIE, t. 12527, 2023, p. 85-92.

[109] O. Ronneberger, P. Fischer et T. Brox, « U-net : Convolutional networks for biomedical image segmentation », in *International Conference on Medical image computing and computer-assisted intervention*, Springer, 2015, p. 234-241.

[110] C. Li et M. Wand, « Precomputed real-time texture synthesis with markovian generative adversarial networks », in *European conference on computer vision*, Springer, 2016, p. 702-716.

[111] F. Samaria et A. Harter, « Parameterisation of a stochastic model for human face identification », in *Proceedings of 1994 IEEE Workshop on Applications of Computer Vision*, `https://cam-orl.co.uk/facedatabase.html`, 1994, p. 138-142. doi : `10.1109/ACV.1994.341300`.

[112] Y. LeCun, C. Cortes et C. Burges, *MNIST handwritten digit database*, 2010.

[113] S. M. R. Hashemi et M. Faridpour, « Evaluation of the algorithms of face identification », in *2015 2nd International Conference on Knowledge-Based Engineering and Innovation (KBEI)*, IEEE, 2015, p. 1049-1052.

[114] A. Ali-Gombe, E. Elyan, Y. Savoye et C. Jayne, « Few-shot classifier GAN », in *2018 International Joint Conference on Neural Networks (IJCNN)*, IEEE, 2018, p. 1-8.

[115] Y. LeCun, L. Bottou, Y. Bengio et P. Haffner, « Gradient-based learning applied to document recognition », *Proceedings of the IEEE*, t. 86, *11*, p. 2278-2324, 1998.

[116] T. M. Damico, « A brief history of cryptography », *Inquiries Journal*, t. 1, *11*, 2009.

[117] F. Cohen, *A short history of cryptography. Retrieved May 4, 2009*, 2007.

[118] D. Kahn, « On the origin of polyalphabetic substitution », *Isis*, t. 71, *1*, p. 122-127, 1980.

[119] A. R. Miller, « The cryptographic mathematics of enigma », *Cryptologia*, t. 19, *1*, p. 65-80, 1995.

[120] M. A. S. Hassan et I. S. I. Abuhaiba, « Image encryption using differential evolution approach in frequency domain », *arXiv preprint arXiv :1103.5783*, 2011.

[121] M. Khan et T. Shah, « A literature review on image encryption techniques », *3D Research*, t. 5, p. 1-25, 2014.

[122] S. A. Khayam, « The discrete cosine transform (DCT) : theory and application », *Michigan State University*, t. 114, *1*, p. 31, 2003.

[123] R. El Saj, E. S. Gooya, A. Alfalou et M. Khalil, « Face recognition and access control applications of the generative classifier pix2pix », in *Pattern Recognition and Tracking XXXIV*, SPIE, t. 12527, 2023, p. 60-66.

COLLEGE MATHSTIC
DOCTORAL BRETAGNE
BRETAGNE OCEANE

ISEN
ALL IS DIGITAL!
yncréa

Université Libanaise
Ecole Doctorale
Sciences et Technologie

UBO
Université de Bretagne Occidentale

Titre : Traitement dans le domaine chiffré par l'intelligence artificielle

Mots clés : Intelligence artificielle, classification par génération, cryptage DCT, pix2pix, auto-encodeur

Résumé Récemment l'intelligence artificielle s'est répandue partout, résolvant une variété de problèmes dans différents domaines et applications. Un système d'intelligence artificielle a besoin d'une grande quantité de données pour s'entraîner, or la disponibilité de ces données représente un défi majeur, surtout que certaines applications traitent de données privées sensibles qui ne doivent pas être partagées ou diffusées. Par conséquent, le développement de systèmes d'intelligence artificielle qui préservent la confidentialité et la sécurité des données devient une nécessité. Le besoin de sécurité est bien plus large que la préservation de la sécurité des données. Toutefois, avec la propagation de la numérisation, il était tout aussi important de développer des solutions qui préservent la sécurité du système en contrôlant l'accès des différents utilisateurs, en donnant l'accès à certains et en le bloquant pour les inconnus. Dans cette thèse, un système de contrôle d'accès basé sur la reconnaissance faciale qui préserve la confidentialité des données est développé. La confidentialité est préservée en cryptant les données à l'aide d'une nouvelle méthode de cryptage basée sur la transformée en cosinus discrète. Ce système de contrôle d'accès utilise une méthode de classification basée sur la génération de données. Cette approche de classification met en évidence la capacité des réseaux tels que pix2pix et l'auto-encodeur, normalement utilisé pour la régénération de données, à distinguer et à classer les données, qu'elles soient cryptées ou non cryptées.

Title : Processing in the encrypted domain using artificial intelligence

Keywords : Artificial intelligence, classification by generation, DCT encryption, pix2pix, auto-encoder

Abstract : Recently artificial intelligence has become widely used, solving a variety of problems in different fields and applications. Artificial intelligence systems need large amounts of data to be trained. However, the availability of data represents a major challenge, especially as some applications use private sensitive data that must not be shared or distributed. Consequently, the development of privacy and security-preserving artificial intelligence systems is becoming a necessity. The security need is much broader than just preserving data security. However, with the spread of digitalization, it was also important to develop solutions that preserve the security of the system by controlling access of different users, giving access to some and blocking it for the unknown. In this thesis, a privacy preserving access control system based on facial recognition is developed. Priacy is preserved by encrypting data using a new encryption method based on the discrete cosine transform. This access control system uses a classification method based on the generation of data. This classification approach highlights the ability of networks such as pix2pix and the auto-encoder, normally used for data regeneration, to distinguish and classify data, whether encrypted or unencrypted.