Medical image enhancement using deep learning and tensor factorization techniques

Janka Hatvani

Université Fédérale
Toulouse Midi-Pyrénées

THÈSE

En vue de l'obtention du
DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par l'Université Toulouse 3 - Paul Sabatier

Cotutelle internationale: Pázmány Péter Catholic University - PPKE ITK

Présentée et soutenue par

Janka HATVANI

Le 31 mai 2021

Amélioration des images médicales à l'aide de techniques d'apprentissage profond et de factorisation tensorielle

---

Ecole doctorale : **EDMITT - Ecole Doctorale Mathématiques, Informatique et Télécommunications de Toulouse**

Spécialité : **Informatique et Télécommunications**

Unité de recherche :
IRIT : Institut de Recherche en Informatique de Toulouse

Thèse dirigée par
Adrian BASARAB et Miklos GYONGY

Jury

M. David BRIE, Rapporteur
M. Bruno SIXOU, Rapporteur
Mme Emilie CHOUZENOUX, Examinatrice
M. Denis KOUAMé, Examinateur
M. Zsolt Szabó, Examinateur
M. Adrian BASARAB, Directeur de thèse
M. Miklós GYöNGY, Co-directeur de thèse
M. Árpád CSURGAY, Président

Medical Image Enhancement using Deep Learning and Tensor Factorization Techniques

Pázmány Péter Catholic University
Roska Tamás Doctoral School
of Sciences and Technology

University of Toulouse III
Toulouse Doctoral School of Mathematics,
Informatics and Telecommunication

Janka Hatvani

Supervisors Dr. Miklós Gyöngy
Dr. Adrian Basarab
Prof. Denis Kouamé

A thesis submitted for the degree of *Doctor of Philosophy*

2021

Abstract

Recording higher quality medical images either requires expensive new devices, poses health risks to the patient, or is limited by physical boundaries. Therefore post-processing resolution enhancement is preferred. The degradation model of the recorded images considered in this thesis assumes a blurred, down-sampled, noisy version of the high-resolution object. Generic super-resolution algorithms estimate this object from a single degraded image (SISR) instead of a series of images or multiple modalities. State-of-the-art reconstruction-based minimization techniques are computationally efficient for 2D images, but not for 3D medical images. In this thesis new algorithms applied to the 3D SISR problem are proposed. The efficiency of the techniques is demonstrated on dental computed tomography (CT) images of extracted teeth. In endodontic treatments the higher resolution CT image is necessary for precise mapping of the dental cavity.

First deep learning was investigated as a possible SISR technique. A subpixel network and the U-net have been considered and applied successfully for the slice-by-slice resolution enhancement of the dental CT image volume.

In the second group of proposed algorithms tensor-decomposition techniques gave the backbone of the 3D SISR frameworks. They are capable of conserving 3D information in a lightweight algorithm structure, while remarkably speeding up the enhancement. Canonical polyadic and Tucker decompositions were investigated, and a semi-blind system-parameter estimation was also implemented.

Compared to the state of the art, the results of both methods suggested better detection of medically salient features such as the size, shape, or curvature of the root canal. The use of deep learning and tensor decompositions in medical 3D SISR was a pioneering work, opening up new perspectives in the field.

Acknowledgements

This thesis was facilitated by the co-tutelle agreement between Pázmány Péter Catholic University of Budapest and Paul Sabatier University of Toulouse – in particular with Institut de Recherche en Informatique de Toulouse (IRIT).

Knowing and feeling that the people around you support, encourage and believe in you, gives great confidence and courage to accomplish your goals. I am lucky to know and work with such people, whom I would like to say thank you.

I have been working with Miklós for six years – hopefully, still counting. I am grateful for the opportunities he opened up to me in Oxford, Toulouse and Dermus, but I appreciate even more what I have learned from him. He taught me how to tackle problems bottom-up, that admitting if you do not understand something is actually a strength, and how to politely answer e-mails which actually infuriate you.

I was very excited for the opportunity of working together with Denis and Adrian on this thesis, and I hope our collaboration will continue. Their enthusiasm and spirit was an integral part of this work, along with the knowledge I was lucky to acquire from them. They make us, their students feel like our problem was the most important of all, and I will always admire their near real-time response to e-mails.

I had the advantage of being part of two teams of doctoral students. In Toulouse Mohamad was my bastion, who helped me with the French bureaucracy, and acquainted me with the Lebanese cuisine, and became a good friend to me. I consider Ákos, Krisztián and Petya also my friends. They always readily help settling any difficulties and make me feel needed when they ask for my help. Thank you all.

Finally I would like to thank my family. My parents always gave me the freedom of choice, and supported me on whatever path I took. Edi, my husband is always encouraging me and treats my work equivalent (if not superior) to his own. I am grateful for your love and help. Sorry for disturbing your sleep with the typing...

Contents

List of Figures

List of Tables

Notations

For easier distinction, scalars are denoted by Greek letters or italic lowercase letters (e.g. α or a), one-dimensional arrays are denoted by bold lowercase letters (e.g. $\mathbf{a}$), 2-dimensional matrices by uppercase letters (e.g. A) and three-dimensional tensors by bold uppercase letters (e.g. $\boldsymbol{A}$). Any of the above with an overline denotes a set of the corresponding elements (e.g. $\overline{A} = \{A_1, A_2, A_3\}$). Technical terms, where they first appear, are *italicized.*

Glossary

Notation	Description	Page
$\circ$	Hadamard product, element-wise multiplication	67
$\oslash$	Hadamard, element-wise division	38
$*$	convolution	16
$\odot$	Khatri-Rao product	30
$\times_n$	mode product along dimension n	28
$\otimes$	outer product of two vectors	27
∂_d	partial derivative in the depth direction	18
∂_h	partial derivative in the horizontal direction	17
∂_v	partial derivative in the vertical direction	17
$\downarrow$	down-sampling operator	16
$\|\cdot\|_1$	ℓ_1-norm	17
$\|\cdot\|_2$	ℓ_2-norm	17
$\|\cdot\|_F$	Frobenius-norm	31
$\|\cdot\|_*$	nuclear norm	18
$\|\cdot\|_{TV}$	total variation norm	17
†	pseudoinverse of matrix	31
vec	lexicographical vectorization of matrix elements	10

Notation	Description	Page
$[\![U_1, U_2, U_3]\!]$	shorthand for composing a tensor from its CPD	30
$[\![\Sigma; V_1, V_2, V_3]\!]$	shorthand for composing a tensor from its TD	33
2D	two-dimensional	2
3D	three-dimensional	2
α	leakage parameter of LReLU	21
ADMM	alternating direction method of multipliers	18
ALS	alternating least squares	30
BCCB	block circulant matrix with circulant blocks	17
BP	back-projection	9
CBCT	cone beam computed tomography	7
CNN	convolutional neural network	19
CPD	canonical polyadic decomposition	29
CT	computed tomography	1
D	decimation matrix	17
dt	update rate of gradient descent	19
ℓ_1	norm and loss of absolute values	17
ℓ_2	norm and loss of squared values	17
ℓ_{1s}	modified ℓ_1 loss	22
η	learning rate of a neural network	22
F	tensor rank	29
FBP	filtered back-projection	9
fMRI	functional magnetic resonance imaging	2
$\mathcal{F}$	Fourier transform	38
G	ground truth output image of neural network	20
H	BCCB Gaussian convolution matrix	17
$\boldsymbol{H}$	Gaussian point spread function tensor	16
HR	high-resolution	14

Notation	Description	Page
I	radiation intensity	8
I_0	radiation source intensity	8
IFC	information fidelity criterion	39
iRT	inverse Radon-transform	9
kVp	kilovoltage peak [kV]	6
ln	natural logarithm	8
LP	line pairs	12
LR	low-resolution	14
LReLU	leaky rectified linear unit	21
LRTV	3D SISR algorithm with low rank and TV regularization	18
µCT	micro computed tomography	8
MRI	magnetic resonance imaging	1
MSE	mean squared error	39
MTF	modulation transfer function	12
N	Gaussian noise matrix	17
$\boldsymbol{N}$	Gaussian noise tensor	16
n_{ADMM}	number of iterations in ADMM	19
NEDI	new edge directed interpolation	15
$\mathcal{N}$	the mapping function of a network	20
n_{grad}	number of iterations in gradient descent	19
NQM	noise quality measure	40
n-rank	{1-rank,2-rank,3-rank} matrix ranks for the n-way matricized tensor	32
O	output image matrix of neural network	20
P	2D projection data, sinogram	8
ϕ	regularization function	17
PET	positron emission tomography	2

Notation	Description	Page
SPECT	single photon emission computed tomography	2
PSF	point spread function	16
PSNR	peak signal-to-noise ratio	39
r	down-sampling rate	43
R	component number of CPD	31
ρ	penalty term of ADMM	19
ReLU	rectified linear unit	21
R_n	component numbers of TD	32
RT	Radon-transform	8
$\overline{\sigma}$	$\{\sigma_1, \sigma_2, \sigma_3\}$ standard deviations along the three dimensions	38
SISR	single image super-resolution	14
SNR	signal-to-noise ratio	9
SR	super-resolution	14
SRR	reconstruction-based super-resolution algorithm	45
SSI	structural similarity index	39
SVD	singular value decomposition	29
SV_n	n-way singular values from the tensor TD	33
T	training input image of neural network	20
TD	Tucker decomposition	32
$\overline{TE}_H$	test set of HR image slices	37
$\overline{\boldsymbol{TE}}_{\boldsymbol{H}}$	test set of HR image volumes	37
$\overline{TE}_L$	test set of LR image slices	37
$\overline{\boldsymbol{TE}}_{\boldsymbol{L}}$	test set of LR image volumes	37
$\overline{\xi}$	neural network parameters (weights and biases)	20
$\overline{TR}_H$	training set of HR image slices	37
$\overline{\boldsymbol{TR}}_{\boldsymbol{H}}$	training set of HR image volumes	37
$\overline{TR}_L$	training set of LR image slices	37

Notation	Description	Page
$\overline{\boldsymbol{TR}_{\boldsymbol{L}}}$	training set of LR image volumes	37
TV	total variation	17
W	projection operator, system matrix	10
W	wavelet transformation matrix	17
X	high-resolution image matrix, 2D attenuation function	8
$\boldsymbol{X}$	high-resolution image tensor	16
$\hat{\boldsymbol{X}}$	estimated high-resolution image tensor	18
$\boldsymbol{X}^{(n)}$	image tensor matricized along direction n	18
Y	low-resolution image matrix	17
$\boldsymbol{Y}$	low-resolution image tensor	16

Declaration of original work used

This section introduces the main references of the current thesis. The main contributions are the journal papers [J1, J2], and conference papers [C1, C2], while [Au1, Au2, Au3, Au4] are internal works, published for the proceedings conference of the doctoral school.

In Chapter 2 Section 2.1.2 the introduction of tomography reconstruction techniques is based on [Au1, Au2]. Section 2.1.4 introduces the main application of the proposed algorithms, dental computed tomography, based on [J2]. Section 2.2 summarizes state-of-the-art resolution enhancement techniques, as previously written in [J1],[Au3]. Section 2.3 uses [J2],[Au4] for presenting the basics of tensor algebra. Chapter 3 is the first contribution of this thesis, based on [J1]. Chapter 4 introduces the second group of contributions, based on multiple articles – [J2] in Section 4.2, [C1] in Section 4.3, and [C2] in Section 4.4.

Further references, which are mostly independent of the author, are provided in the appropriate (sub)sections. To facilitate differentiation, modified/adapted or images taken directly from other sources are cited at the end of the figure caption, whereas pictures created by the author are not cited.

Chapter 1

Introduction

1.1 Motivation

For precise medical diagnosis the doctor often needs to see inside the body to have a better understanding of the underlying bodily state. For avoiding the complications of surgical interventions, non-invasive medical imaging techniques (referred to as modalities) are preferred. Depending on the tissue type, body function, or life processes one would like to image there are multiple imaging techniques at the disposal of healthcare providers.

Ultrasound imaging is a cost-effective and safe technique, where the resolution and penetration depth are inversely proportional – while high-frequency images of the skin have excellent resolution (<0.5 mm [1, p. 23]) , the quality of fetal scans or cardio-vascular flowmetry is moderate (<3 mm [1, p. 23]).

Conventional X-ray and *computed tomography* (CT) imaging provide important information, especially in the case of air pockets and hard tissue. It is widely used both in preventive and diagnostic imaging (with a resolution of $\tilde{1}$ mm [1, p. 23]) for a relatively low price. This modality is ionizing, therefore should be used only after considering the possible health risks. An increased radiation dose or higher number of scans give better quality images, but also aggravates its hazardous effects.

Magnetic resonance imaging (MRI) is a safe modality when used in a secure environment, but is relatively expensive. Its long scanning time makes this technique sensitive to motion artefacts, posing an additional challenge to radiologists. It gives

detailed images (resolution $\tilde{1}$ mm [1, p. 23]) of soft tissues rich in water molecules, where CT usually fails. It can also be used for following bodily functions, namely brain activity (fMRI).

Positron emission tomography (PET) and single photon emission computed tomography (SPECT) give an insight on the metabolism of the body using radioactive substances. They have relatively bad resolution (>2 mm [2]) but give information on bodily functions, which cannot be provided by other modalities. These techniques demand the highest operation cost and has to be used in combination with CT or MRI imaging to register the function with the anatomy.

As explained above, all imaging modalities have a boundary on their achievable resolution, either caused by a physical limit, health risks, or imaging artifacts. As a consequence, upgrading the hardware to obtain a better resolution is either expensive, might imperil the patient or is physically not possible at all. Post-processing resolution enhancement algorithms offer a solution to these problems. A widespread and easy technique is interpolation, providing limited enhancement. Reconstruction-based minimization techniques are the state of the art in terms of image quality, providing a computationally efficient method in the case of two-dimensional (2D) images. However, most medical images are three-dimensional (3D) (or have even higher dimensionality), where state-of-the-art techniques fail to give computationally efficient results.

The aim of this thesis is to offer new resolution enhancement methods, particularly for 3D medical images using a single scan. In particular, deep learning, and tensor implementations of the reconstruction-based minimization is investigated. The algorithms presented may be used in different modalities, however, they are validated through a dental CT application. Root canal treatment is a common procedure in dentistry, which requires the precise mapping of the 3D canal structure. Currently used dental CT imaging does not always provide sufficient resolution for the treatment. Increased radiation dose gives more detailed scans, but is potentially carcinogenic, making post-processing applications good candidates for enhancing dental image quality.

In the light of the above, the central research question investigated are:

1. Is deep learning a viable method for dental CT single image super-resolution?
2. How is tensor factorization applicable in 3D single image super-resolution?
3. Do tensor implementations of the 3D single image super-resolution problem offer faster algorithms than the current state of the art does?
4. Can the system parameters be estimated within a tensor framework of the 3D single image super-resolution problem?

1.2 Overview of the thesis

This section provides a brief overview of the current thesis.

Chapter 2, is an overview of the background to the topic and underlying theory. As the main scope of the thesis is CT imaging, Section 2.1 reviews the physics, image reconstruction and quality, and endodontic use of this technique. Section 2.2 presents common resolution enhancement methods, such as interpolation, deconvolution-based minimization, and deep learning. Finally, Section 2.3 introduces the basics of tensor algebra and decompositions, providing the basis of the proposed algorithms in Chapter 4.

Chapter 3 is the first contribution of the current thesis. The topics of Section 3.1 are the dental CT images, estimated system impulse response and applied metrics, which were used for testing the proposed algorithms. The rest of the chapter introduces a deep learning algorithm for the resolution enhancement of the CT slices, together with its novel results.

The second group of proposed algorithms are listed in Chapter 4. Section 4.1 introduces the image degradation model of the dental CT images. The first algorithm using canonical polyadic decomposition is described in Section 4.2, which is extended with a semi-blind system function estimation in Section 4.3. In Section 4.4, Tucker-decomposition is used for the resolution enhancement.

Chapter 5 answers the posed research questions in Section 5.1 and elaborates on the opened perspectives of the introduced work in Section 5.2.

Finally, in Chapter 6 the new scientific results are summarized in the form of thesis points in English, French and Hungarian.

Chapter 2

Background and Theory

2.1 Computed tomography imaging

CT is a wide-spread diagnostic imaging tool that uses 2D X-ray projections to reconstruct a 3D map of the scanned object. It is used for, without any claim to completeness, detecting bone trauma, fractures, inflammations, tissue death after stroke, tumors, lesions of the lung, or the state of the cardio-vascular system through angiography. This large set of applications makes it one of the most frequently used medical imaging modalities. In Hungary 122.5 (in 2017), in France 189.7 (in 2017),

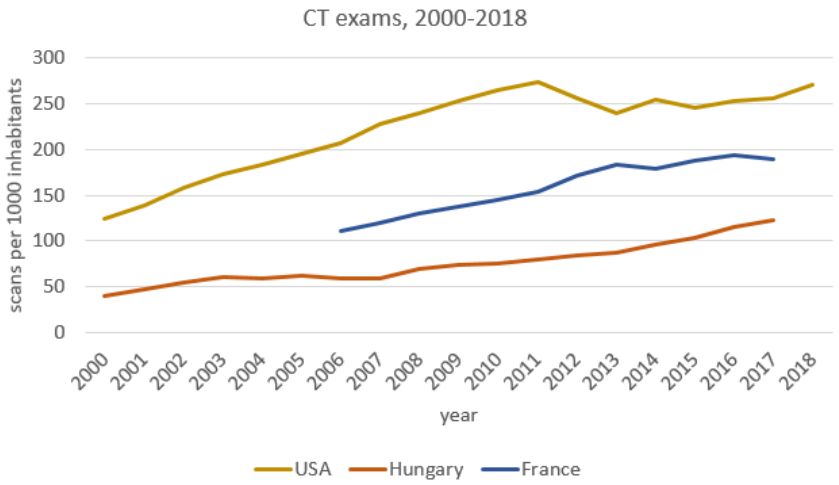

Figure 2.1: Number of CT scans acquired per 1000 inhabitants in France, Hungary and the United States. The increasing number of scans is controversial, as apart from granting a more precise diagnosis, the radiation also poses a health risk. The data was taken from [3].

in the United States 271 (in 2018) scans were acquired per 1000 inhabitants [3]. The growing number of exams in the three countries can be seen in Fig. 2.1. However, the increased radiation dose in these patients increases the carcinogenic potential [4]. This is especially true for children, where CT scans have been found to triple the risk of developing leukemia, brain tumor and bone marrow cancer [5]. Thus, decreasing the dose of the ionizing radiation would be beneficial for making this otherwise extremely useful diagnostic tool safer.

2.1.1 Physics of CT imaging

In a CT machine the X-ray beam is generated in a vacuum tube. It penetrates the object to be examined. As it interacts with the atoms of the examined object, it will be partially absorbed and scattered. The *X-ray photons* (photon with a frequency in the range 30×10^{15} Hz - 30×10^{18} Hz) exiting the object will be collected on a detector grid, resulting in a scan. As the generator-detector pair rotates around the object, multiple scans will be recorded, which will be used for the reconstruction of the 3D volume.

Generation of X-rays

In the X-ray vacuum tube, a cathode emits, an anode collects electrons, resulting in a flow of electrical current, the beam. A characteristic of the system is the *kilovoltage peak* (kVp, [kV]), the amplitude of the potential connected to the anode-cathode pair, accelerating the flow of electrons. As these accelerated electrons collide with the anode, they will be decelerated. The difference in kinetic energy will be turned into 99% heat and 1% X-ray photons, called Bremsstrahlung (breaking radiation), perpendicular to the direction of the beam. This effect is presented as a continuous line on the radiation spectrum. Further X-ray photons are generated when the high-speed electrons collide with atomic inner-shell electrons. The excited atomic electron leaves the shell and a higher-energy outer-shell electron fills the vacancy. The energy difference between the two shells will leave the atom in the form of an X-ray photon. The wavelength of these photons is characteristic to the material it is fabricated from (usually tungsten, molybdenum or copper), adding

sharp peaks to the spectrum [6].

Interaction of X-rays with tissue

Photoelectric absorption of the X-ray photons happens similarly to their generation [7, Section 3.3]. As they collide with atoms of the examined object, they may transfer their energy to orbital electrons, causing their ejection from the hosting atom. This electron will ionize neighboring atoms. The vacancy left on the hosting atom will start a chain reaction, as it will be filled by a higher-energy electron from outer shells, generating a new X-ray photon, colliding with a lower-energy electron, until the excess energy dissipates. When Compton (incoherent) scattering happens, the photon changes direction upon colliding with an electron, transferring part of its energy to the electron, depending on the scattering angle. This scattered photon may hit the detector from odd angles, leading to contradiction in the reconstruction.

Deterministic effects of radiation happen above a threshold intensity in every patient, within hours or days. Such consequences include skin irritation, burn, sterility, damage of the eye lenses, or even death. Stochastic effects cannot be linked to a standard threshold, risk increases with dose, and develop over years after the exposure. Genetic modifications and cancer are such disorders.

CT machine designs

In the first CT machines a single detector element collected a small portion of the beam, approximating a single line of projection. The source-detector element was translated over time along a line, resulting in a parallel sampling geometry (Fig. 2.2 a). This relatively slow design was subject to movement artifacts, so engineers came up with new concepts. In the fan-beam geometry detector elements are positioned along an arc, recording multiple rays emitted at the same instant (Fig. 2.2 b). *Cone beam CT* (CBCT) uses an either flat or curved 2D detector array, utilizing a larger portion of the beam (Fig. 2.2 c). This reduces motion artifacts and radiation dose, but requires more complex reconstruction algorithms.

In CBCT usually a single rotation of 200° is covered. For dose reduction a smaller range or sparser sampling might be applied, however, this results in heavy

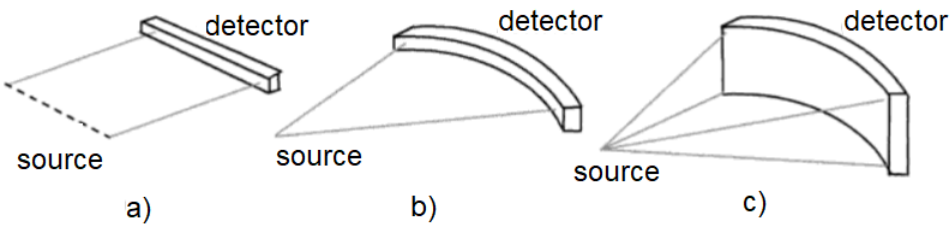

Figure 2.2: Sampling geometries used in CT. a) Parallel beam, b) fan beam, c) cone beam. Image source: [8, pp. 44].

artifacts. With the fan-beam geometry the projection set might be acquired in a step-and-shoot manner, where a predefined number of slice scans are taken. In the helical acquisition used with fan-beam the patient or source-detector pair is moved continuously, resulting in a spiral trajectory. This allows fast scanning with reduced motion artifacts, but requires more complex reconstruction techniques. In multi-slice CT multiple (2-64) detector arcs are installed, so multiple scans are recorded under one rotation, offering better resolution under the same radiation level, but for a higher hardware cost.

Medical CT machines use relatively low radiation doses to limit the ionizing effect. *Micro-CT* (µCT) machines operate on a higher radiation dose to give a well-detailed map of small objects, and can also be useful in medical *ex vivo* applications.

2.1.2 Reconstruction of the 3D volume

While in X-ray imaging the detector might be a film, in CT machines it is a digital grid, allowing the computational reconstruction of the 3D volume. The mathematical descriptor of the imaging process is the *Radon-transform* (RT) or forward projection, which in 2D form is [9]

$$P(\theta,t) = -\ln\left(\frac{I(\theta,t)}{I_o}\right) = \mathrm{RT}(X) = \int_{l_{\theta,t}} X(x,y)\mathrm{d}s(x,y), \tag{2.1}$$

where $P(\theta,t)$ is the projection data (sinogram) acquired under angle θ at coordinate t on the projection axis. $I(\theta,t)$ is the intensity measured at the same point, I_0 is the source intensity. The line integral of the attenuation function X along the projection line equals the value at $P(\theta,t)$. The ray intensity is attenuated along this

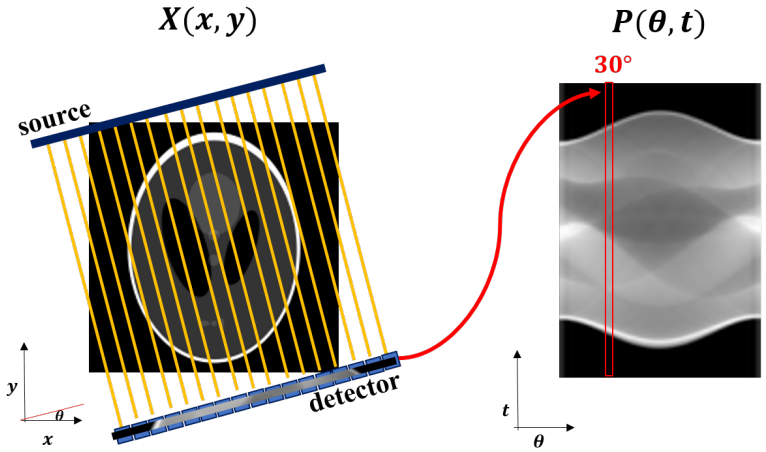

Figure 2.3: The Radon-transform. The attenuation function $X(x, y)$ is transformed to the sinogram domain, resulting in $P(\theta, t)$. θ is the projection angle and t is the detector element (this notation originates from the translation of a single detector element in Fig. 2.2 a). Thus the line integral along a yellow ray (θ, t) of $X(x, y)$ is calculated for a point of $P(\theta, t)$.

line from the source, which is compensated by taking the logarithm of the normalized measured intensity. The RT is illustrated in Fig. 2.3.

The basic concept of direct reconstruction methods is to simply *back-project* (BP) the measured values onto the imaged volume:

$$X(x, y) = \int_0^{\pi} P(\theta, x \cdot \cos(\theta), y \cdot \sin(\theta)) \mathrm{d}\theta. \tag{2.2}$$

In the inverse Radon-transform (iRT) the backprojection is accompanied by a ramp filtering step (*filtered backprojection*, FBP) - it arises from the coordinate transformation $(x, y) \rightarrow (\theta, t)$. This filter is prone to intensifying high-frequency noise [10], and is in practice often combined with a Hamming, Hanning, Cosine or the Shepp-Logan filter. In 3D reconstruction usually the 2D iRT is used to reconstruct planes of the 3D object. The more complex 3D RT requires rotations around two angles for a full data-set, and plane integrals are considered.

Incomplete information or low *signal-to-noise ratio* SNR makes iRT unstable. In this case iterative techniques offer better solution for the reconstruction, which is demonstrated on the Shepp-Logan phantom in Fig. 2.4.

In iterative algebraic reconstructions the system is assumed to be fully numerical, considering discrete pixels and projections. The RT is written in the form of a system

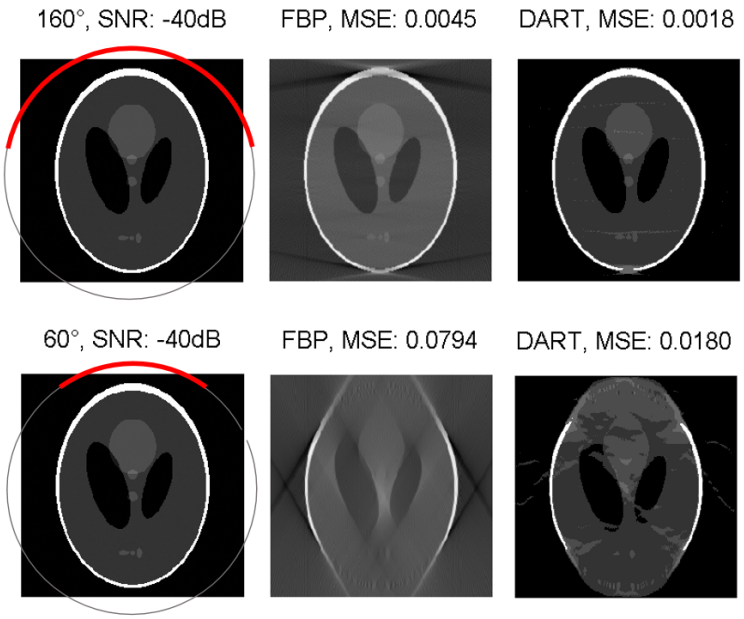

Figure 2.4: FBP becomes unstable under noise or limited information. The upper row shows analytic FBP and iterative DART reconstructions from projections taken under 160° (red arc), the lower row under 60°. 40 dB noise was added to the projection set. The error of the reconstructions is printed above the images.

of linear equations, which in matrix-vector form is

$$\mathrm{vec}(P) = W \cdot \mathrm{vec}(X), \tag{2.3}$$

where $\mathrm{vec}(P)$ is the lexicographical *vectorization* of $N \times 1$ measured values on the detector elements, $\mathrm{vec}(X)$ is the discretized and vectorized function of attenuation values of size $M \times 1$, and W is an $M \times N$ matrix where element W_{ij} represents the contribution of j^{th} image-pixel to the i^{th} detector element [Au1, Au2]. This matrix W is called the projection operator or system matrix [9]. The task is to recover X from the known projection data P and system matrix W. The algorithms iteratively update the unknown coefficients in X by minimizing the difference between the measured and forward projected detector values.

The strength of the iterative methods over analytic techniques is their flexibility, making prior knowledge easy to be incorporated into the algorithm. Such an information can be the non-negativity of the image, as no negative attenuation will be present in the reconstruction field. Another method is to predefine the number of possible discrete attenuation levels, as in the discrete algebraic reconstruction

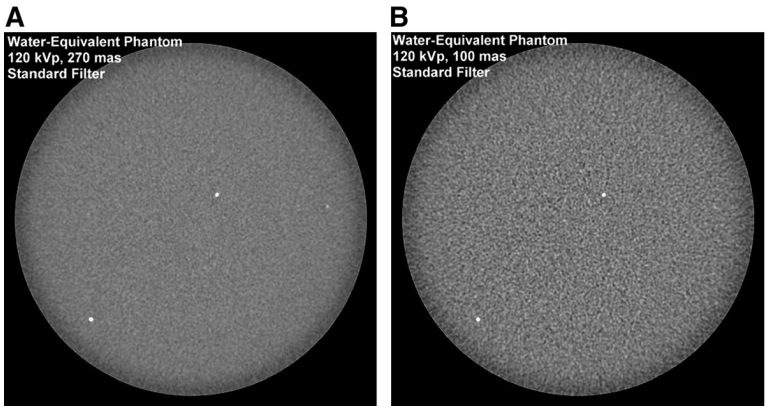

Figure 2.5: CT image noise from 270 mAs and 100 mAs scans of a phantom. The lower tube current (hence radiation dose) causes a lower SNR. Image was taken from [12]

technique (DART).

2.1.3 X-ray dose and image quality

The *absorbed dose* is the radiation energy (joule, [J]) deposited over a unit mass of tissue [kg], and its measure is gray [Gy]. The absorbed dose is usually weighted by tissue-specific constants in order to find the equivalent dose, measured in sievert [Sv], indicating the likely deterministic and stochastic biological effects.

Both definitions of dose are calculated using the radiation energy, product of the tube current [mA], exposure time [s] and kVp [kV]. For achieving a good penetration and (SNR), a high number and high intensity of X-ray photons is necessary. The number of photons is proportional to the tube current and kVp, while the intensity depends on the kVp. For a constant detector exposure and radiation dose the kVp and tube current need to be changed inversely. However, high energy photons are more likely to be scattered than absorbed, which effect is less tissue-dependent, leading to low-contrast images [11]. In Fig. 2.5 the contrast is kept constant by only decreasing the tube current, leading to lower radiation dose and higher noise level.

The diagnosis depends heavily on the quality of the CT image, and different clinical tasks need different SNR. Detecting a solid nodule in the lung or calcium in the coronary artery requires a high image quality, while abdominal or diffuse lung diseases are usually detectable under low SNR [13]. This resolution is limited by diffraction, aberration, the focal spot size of the beam, the detector size, the

slice thickness, noise, the reconstruction algorithm, the applied filters, or patient movement [14].

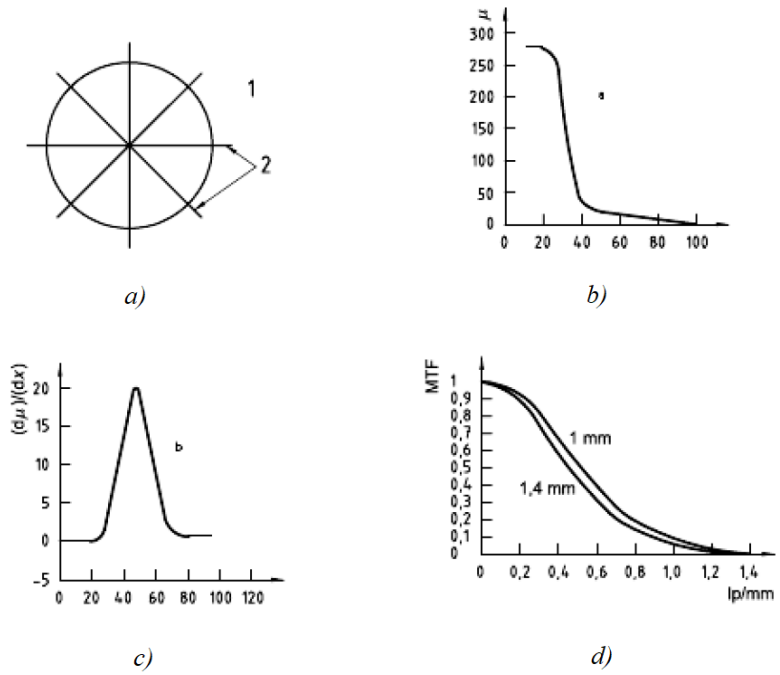

Figure 2.6: Calculation of the MTF, using a cylinder phantom. In a) *one-dimensional profiles are taken through the center of the cylinder, in* b) *these aligned profiles are averaged. In* c) *the average profile is differentiated, and its Fourier transform, the MTF is calculated in* d), shown for two different aperture settings. *Image source: [14].*

For measuring the resolution of a radiological system two conventions are used in the standards. The *modulation transfer function* (MTF) is measured on the edge of a cylinder as explained in Fig. 2.6. The MTF is normalized to one, and the spatial frequency is given in *line pairs over mm* (LP/mm). This provides the contrast amplitude as a function of spatial resolution. The second approach measures the contrast of alternating line pair structures against air, as depicted in Fig. 2.7. The contrast level at different LP/mm is directly calculated. Manufacturers usually express the resolution of the system as [n LP/mm at m% MTF]. Following the above description it means that n line pairs stacked within 1 mm will be distinguishable at m % contrast level.

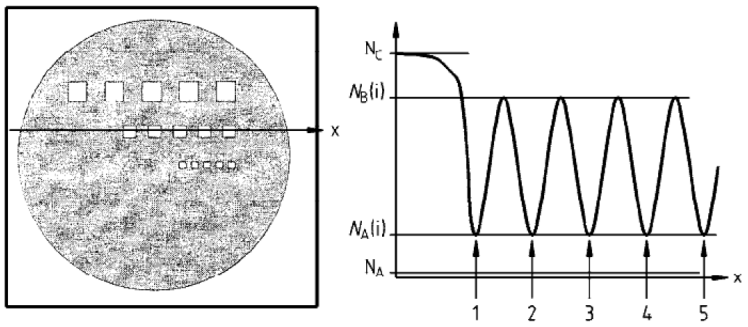

Figure 2.7: Calculation of the spatial resolution from air-material line pair structures. The gray profile is extracted along the structure lines. The contrast factor is calculated from the average difference of its minima (at cut-outs, air), and maxima (material bridges), normalized by the amplitude difference of undisturbed material N_A and undisturbed air N_C. Image source: [14].

2.1.4 Relevance of resolution enhancement in dental CT

Endodontics is the dental specialty concerned with the maintenance and treatment of the dental pulp, formed by nerves, blood vessels and connective tissues within the root canal. According to [15], it is an indispensable prerequisite to identify the root canal before the treatment, so tools of adequate length and diameter can be chosen, allowing precise disinfection and obturation (hermetic sealing by filling empty spaces) of the cavity. Even though endodontic treatment is one of the most common procedures, epidemiological studies show success rates of only 60-85% for general practice [16, 17]. Consequences are periapical diseases and their consequences on health, such as the future of the treated teeth, the prosthetic replacement of the extracted tooth on the jaw, or the impact on cardiovascular and diabetic diseases. The reduction of endodontic therapeutic failures require new techniques for improving the quality of endodontic treatments [18–22]. In dentistry the 3D structure of the tooth is visualized using CBCT, where the typical resolution is around 500 µm [23]. When the exact position of the dental canal has to be determined for root canal treatment, these images are difficult to work with, since the diameter of the canal is usually in the range of 0.16-1.6 mm [24]. The European Commission on Radiation Protection concluded in 2012 that further research to establish the diagnostic accuracy of dental CBCT devices in identifying root canal anatomy is necessary to justify their indication in endodontic treatment [25]. *Ex vivo* µCT

gives a sufficient resolution for precise segmentation of the pulp cavity but can be used only on extracted teeth due to size limitations of the hardware, long acquisition time, and high radiation dose. An algorithmic solution for approximating the resolution of µCT images from CBCT acquisitions would therefore be advantageous.

2.2 Resolution enhancement techniques

The resolution enhancement may happen either on the hardware or on the software side. In CT a naive way to improve the image quality in terms of reduced noise could therefore be to increase the radiation dose, which would, on the other hand, also increase the hazardous effects on health and negatively influence the contrast level.

Post-processing has the advantage of increasing the image quality without changing the hardware. The output *high-resolution* (HR) image is estimated from input *low-resolution* (LR) images(s). Widespread *interpolation* methods are merely the upsampling of the image, only handling supposed aliasing effects caused by a *downsampling* of the detector grid. On the other hand, *super-resolution* (SR) algorithms aim to mitigate the problem of downsampling, *blurring*, and added noise introduced during recording [Au3]. SR algorithms can be sorted into three groups depending on the input. This input can be a series of LR images, as in the case of 4D CT, where a sequence of scans is used, compensating for the respiratory movement artifact [26]. In the second group images from different modalities are used, like the slightly different acquisitions of spectroscopic (multi-energy) CT [27]. However, the most common application uses one single capture of the 3D volume (the enhancement named hence *single image super-resolution*, SISR), marking SISR as the scope of this thesis [28–30], [J1, J2]. Most SISR algorithms are implemented for 2D applications. 3D images, like a CT-scan, pose additional difficulties, as these large image files require increased computational capacity in terms of memory and time. For this reason the implementation of 2D algorithms for 3D applications demand new calculation methods.

This section gives an overview of 3D resolution enhancement techniques, starting

with the interpolation techniques, followed by state-of-the-art reconstruction-based SR techniques, and finally introducing deep learning solutions, supporting the first contribution of this thesis.

2.2.1 Interpolation

Interpolation methods are relatively cost-efficient and data-independent solutions for resolution enhancement, but only upsample the image. It is commonly a step of more advanced SR algorithms, which methods also deal with added noise and blurring. During interpolation unknown data points of the new grid are estimated as values of a function fitted to the known samples.

In the *nearest neighbor method* each unknown grid point is assigned the value of the closest known sample. This results in piece-wise constant areas around the known samples, giving the impression of larger image pixels.

The *linear interpolation* in one dimension fits a straight line to the two neighboring known samples, and evaluates this function at the unknown grid point. For higher dimensions the same steps are repeated in the remaining directions.

The one-dimensional *cubic interpolation* fits a third degree polynomial on four neighbouring known samples, and evaluates this function in the unknown grid point. For higher dimensions it is also repeated along the remaining directions. This method is commonly used in many applications, as it gives sharper outputs compared to nearest neighbor and linear interpolation, and is still relatively computationally efficient.

Other functions, as higher order polynomials, splines and sinc retain more information, but are computationally more demanding. All these techniques suffer from artifacts such as aliasing (visible on the edges of Fig. 2.8 b), blurring (see the edges and low contrast of Fig. 2.8 c), or edge halo (the dark contour around the white edges of Fig. 2.8 d). Adaptive interpolation techniques attempt to eliminate or minimize these artifacts at the cost of more complex computations and adding spurious pixels to the image (single pixels near the edges of Fig. 2.8 e), which may or may not improve the overall quality of the enhancement, depending on the texture. New edge directed interpolation (NEDI) is a commonly used example, which estimates

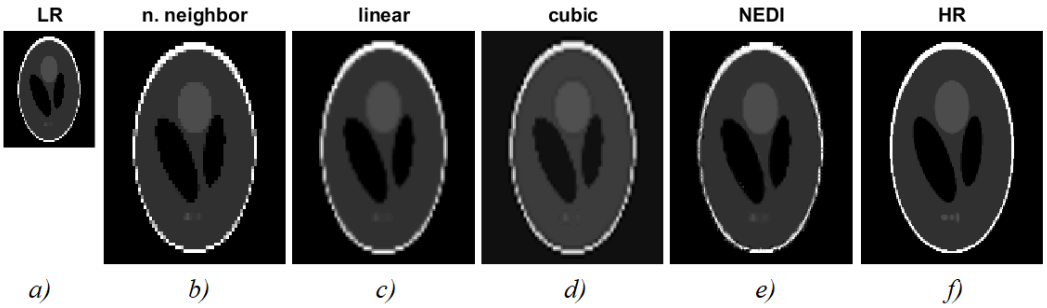

Figure 2.8: Interpolation techniques presented on the Shepp-Logan phantom. Left to right the low resolution image, the nearest neighbors, linear, cubic, edge-directed interpolations and the high resolution image is shown. The NEDI was not implemented for 3D and is evaluated on a single slice, the rest of the algorithms were run on the 3D Shepp-Logan volume.

the image edges form the local covariance and tunes the interpolation coefficients accordingly, preserving the image edges [31].

2.2.2 Reconstruction-based SR methods

Reconstruction-based algorithms are among the most popular SR techniques [29, 30]. In the following the image *degradation model* is defined for 3D images using the *tensor*-notation (generalization of the vector to higher dimensions). These methods assume that the observed LR image $\boldsymbol{Y}$ can be thought of as a blurred, downsampled (denoted by operator $\downarrow \{\}$), and noisy (added Gaussian noise, $\boldsymbol{N}$) version of the HR image $\boldsymbol{X}$. The blurring effect is generally modelled as a convolution with a spatially-invariant point-spread function (PSF, denoted by $\boldsymbol{H}$). With this assumption the degradation model can be written as

$$\boldsymbol{Y} = \downarrow \{\boldsymbol{H} * \boldsymbol{X}\} + \boldsymbol{N}. \tag{2.4}$$

The degradation model has to be inverted in order to obtain $\boldsymbol{X}$. In case the inverse of this function does not exist, is not unique or is not continuous, it yields an *ill-posed* problem. In case of the degradation model, the added noise and the down-sampling make $\boldsymbol{X}$ not directly recoverable, the problem is ill-posed.

For solving ill-posed inverse problems, *regularization* is applied for noise-robust results. The prior information of the regularization might be similar known images — usually difficult to obtain,— or some properties of the image, like sparseness in

a given domain, or possible value ranges. Common regularizers are total variation (TV) promoting piece-wise constant solutions [32], low rank [29], wavelet frames [33] or patch-based structural similarity constraint [34]. Candes et al. have shown that even highly incomplete images can be exactly reconstructed through convex optimization given the sparseness of the image in the gradient domain [35].

The fast SISR algorithm for 2D images

Zhao et al. gave a solution for the 2D SISR problem [30]. The cyclic convolution is written as a multiplication by a *block circulant matrix with circulant blocks* (BCCB, denoted by H), and the downsampling is also realized by a matrix multiplication with D written as

$$\mathrm{vec}(Y) = DH\mathrm{vec}(X) + \mathrm{vec}(N), \tag{2.5}$$

where 2D matrices Y, X and N are lexicographically vectorized. The image X is estimated by solving the minimization problem

$$\min_X \frac{1}{2} \|\mathrm{vec}(Y) - DH\mathrm{vec}(X)\|_2^2 + \tau\phi(A\mathrm{vec}(X)), \tag{2.6}$$

where the first term is establishing data fidelity using the ℓ_2-norm ($\|\cdot\|_2$), and the second imposes some regularization $\phi(\cdot)$ derived from the prior information, as TV, low rank, wavelet transform norm, etc. The constant τ balances the two terms, and A stabilizes the function ϕ. The authors in [30] proposed a fast analytical solution when the regularizer term is the ℓ_2-norm

$$\phi(A\mathrm{vec}(X)) = \|A\mathrm{vec}(X) - v\|_2^2 . \tag{2.7}$$

A general solution aims to solve the problem for different regularizers, as the TV-norm ($\|\cdot\|_{TV}$ with horizontal and vertical derivatives $[\partial_h, \partial_v]$ of the 2D image)

$$\phi(A\mathrm{vec}(X)) = \|X\|_{TV} = \sqrt{\|\partial_h X\|_2^2 + \|\partial_v X\|_2^2}, \tag{2.8}$$

the ℓ_1-norm ($\|\cdot\|_1$) in the wavelet domain (with a transformation matrix W)

$$\phi(A\mathrm{vec}(X)) = \|\mathrm{W}\mathrm{vec}(X)\|_1 , \tag{2.9}$$

or with learning-based regularization, jointly training an LR and an HR patch-dictionary. This general algorithm[1] uses the alternating direction method of multipliers (ADMM) framework. This method turns the problem into several subproblems with variable splitting, where some of the subproblems will be in the form of (2.7), leading to a computationally efficient, fast and smoothly converging solution for 2D images.

Low rank - total variation 3D SISR algorithm

Among the relatively small collection of 3D SISR techniques, LRTV provided competitive results compared to other popular methods (cubic interpolation, non-local means, TV-based up-sampling) [29]. It uses low-rank and total-variation regularizers, minimizing the following cost-function

$$\hat{\boldsymbol{X}} = \underset{\boldsymbol{X}}{\operatorname{argmin}} \left\| \downarrow \{\boldsymbol{H} * \boldsymbol{X}\} - \boldsymbol{Y} \right\|^2 + \lambda_R Rank(\boldsymbol{X}) + \lambda_{TV} \left\| \boldsymbol{X} \right\|_{TV}, \tag{2.10}$$

where λ_R and λ_{TV} are hyperparameters. The $Rank()$ regularizer builds on 2D matrix norms. It sums the trace norms (sum of singular values, denoted by $\|\cdot\|_*$) of the matricized 3D image ($\boldsymbol{X}^{(n)}$, explained later in (2.25)) along each direction ($n \in \{1, 2, 3\}$) as

$$Rank(\boldsymbol{X}) = \sum_{n=1}^{3} \left\| \boldsymbol{X}^{(n)} \right\|_*. \tag{2.11}$$

The low rank condition helps to recover missing values based on global information of known pixels. Similarly to (2.8), $\|\boldsymbol{X}\|_{TV}$ becomes

$$\|\boldsymbol{X}\|_{TV} = \sqrt{\|\partial_h \boldsymbol{X}\|^2 + \|\partial_v \boldsymbol{X}\|^2 + \|\partial_d \boldsymbol{X}\|^2}, \tag{2.12}$$

where ∂_h and ∂_v are the partial derivatives defined for (2.8) and ∂_d is the partial derivative in the depth-direction, the third dimension. This condition helps to keep local information, preserving edges and promoting piece-wise constant patches.

The minimization problem (2.10) is solved in the ADMM framework, by rewriting

[1]The Matlab code associated with fast 2D SISR is available at www.irit.fr/ Adrian.Basarab/codes.html

(2.10) as

$$\min_{\boldsymbol{X},\{\boldsymbol{M_n}\}_{n=1}^3,\{\boldsymbol{L_n}\}_{n=1}^3} \|\downarrow\{\boldsymbol{H}*\boldsymbol{X}\}-\boldsymbol{Y}\|^2+\lambda_R\sum_{n=1}^{3}\left\|\boldsymbol{M_n}^{(n)}\right\|_* + \lambda_{TV}\|\boldsymbol{X}\|_{TV}+\sum_{n=1}^{3}\frac{\rho}{2}(\|\boldsymbol{X}-\boldsymbol{M_n}+\boldsymbol{L_n}\|^2-\|\boldsymbol{L_n}\|^2), \quad (2.13)$$

with the splitting of

$$\boldsymbol{X}=\boldsymbol{M_n} \text{ for } n=\{1,2,3\},$$

where $\{\boldsymbol{L_n}\}_{n=1}^3$ are Lagrangian parameters, and ρ is a penalty term. The problem (2.13) is optimized in n_{ADMM} iterations by minimizing sequentially for $\boldsymbol{X}$ (using gradient descent with iteration number n_{grad} and an update rate dt), $\{\boldsymbol{M_n}\}_{n=1}^3$, and $\{\boldsymbol{L_n}\}_{n=1}^3$ (the latter two having closed-form solutions).

The gradient descent and extensive matrix-operations make this algorithm[2] computationally inefficient, and the many parameters are difficult to tune. However, these two regularizers help to combine local and global information, a successful solution to the higher dimension SISR problem.

2.2.3 Deep learning

Deep neural networks – in particular *convolutional neural networks* (CNNs) – have been shown to be powerful tools in image processing in the last couple of decades. In biomedical imaging CNNs are mainly used for classification, segmentation and detection. Some examples for these kinds of tasks are differential diagnosis between Alzheimer's and Huntington's diseases on MRI data [36], tumor segmentation with multiscale analysis [37], striatum segmentation [38], or tumor and lesion detection [39], classification [40].

While deep learning is increasingly practiced in the above areas of biomedical imaging, its use in image enhancement is less investigated. Deep learning has been used so far for image denoising [41], image generation, as constructing CT images from MRI data [42], or artifact removal from sparse-view [43, 44] and limited-angle

[2]The Matlab code associated with LRTV is available at https://bitbucket.org/fengshi421/superresolutiontoolkit

CT images [45]. Before [J1], biomedical image resolution enhancement with deep learning has only been implemented using multi-input frameworks - the input was either an LR cardiac MRI sequence [46, 47] or multichannel MRI data (T1- , T2-weighted and Fluid Attenuated Inversion Recovery images [48]). Most of the previously mentioned biomedical applications tend to use CNNs [36–40, 42, 46].

In the following the basics of deep learning algorithms will be introduced.

A network of artificial neurons

The base unit of a neural network is an *artificial neuron*, with a workflow shown in Fig. 2.9. These working units can be connected in a network, organized in *layers*. A neuron takes as an input the output of single neurons from the previous layer. Usually the deeper the network, the more complex tasks can be learned.

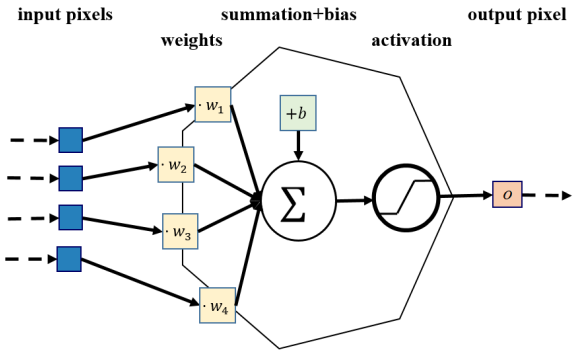

Figure 2.9: The artificial neuron takes the weighted (w_n) sum of its inputs (blue squares), shifts it by a bias (b), and following a non-linear function gives an output (pink square).

The *weights* and *biases* of the network (together denoted by parameters $\overline{\xi}$) are trained using a learning set of known input-output pairs (T for a single training input and G for corresponding ground truth output). During an iteration a network output (O) is generated using the current $\overline{\xi}$, by passing T through the connected neural functions of the network (denoted by $\mathcal{N}$) as

$$O = \mathcal{N}(\overline{\xi}, T). \tag{2.14}$$

The error between O and G is calculated using a selected *loss function*, and is *backpropagated* with a gradient-descent-type algorithm to update the parameters of the network.

Activation function

Each artificial neuron gives its output using a non-linear function, called *activation function*. This function and its derivative should be easy to compute, as they are applied on each node of the network. A common example is the *rectified linear unit* (ReLU), written as

$$\mathrm{ReLU}(x) = \begin{cases} x & \text{if } x \geq 0 \\ 0 & \text{if } x < 0, \end{cases} \tag{2.15}$$

where x is the weighted, biased sum within the neuron. This function is, however, prone to the dying ReLU problem, which happens when a neuron outputs zero, and will keep giving zero response throughout the training. This is caused by the derivative of ReLU, which gives zero gradient for such neural outputs during backpropagation. Leaky implementations of ReLUs have been shown to provide higher accuracy and avoid the dying ReLU problem by providing a non-zero gradient for the constantly inactive neurons in the network [49]. The function of the *leaky-ReLU* (LReLU) is

$$\mathrm{LReLU}(x) = \begin{cases} x & \text{if } x \geq 0 \\ \alpha x & \text{if } x < 0, \end{cases} \tag{2.16}$$

where α is a parameter defining leakage of the ReLU over negative responses, which provides a gradient to compensate for wrongly initialized or trained values. Parameter α is typically a small positive number.

Loss function

The error between the network and expected output is the cost function $J(\cdot, \cdot)$ of the network,

$$J(O, G) = J(\mathcal{N}(\overline{\xi}, T), G). \tag{2.17}$$

For applying the gradient descent-type backpropagation algorithms for updating w and b, this cost (or loss) function needs to be fully differentiable. In the following some possible error functions are listed.

The so-called ℓ_1 loss is the absolute difference,

$$\ell_1(O, G) = |O - G|\,, \tag{2.18}$$

while the ℓ_2 loss is the squared difference (notice the analogy with the ℓ_1- and ℓ_2-norm), written as

$$\ell_2(O, G) = (O - G)^2. \tag{2.19}$$

The ℓ_1 loss is generally better for SR problems as well as for texture and image generation, since ℓ_2 loss is often dominated by outlier pixels on the ground truth images [50]. On the other hand, while ℓ_2 is twice, ℓ_1 is only once differentiable (except for the origin, where $O = G$ and the ℓ_1 function is not differentiable at all).

Huber loss [51] combines the advantages of the ℓ_1 (2.18) and ℓ_2 (2.19) loss functions, helping the network to avoid local minima during training. The twice differentiable and smoother loss function ℓ_{1s} is:

$$\ell_{1\mathrm{s}}(O, G) = \begin{cases} \delta|O - G| - (1 - c)\delta^2 & \text{if } |O - G| > \delta \\ c(O - G)^2 & \text{if } |O - G| \leq \delta, \end{cases} \tag{2.20}$$

where δ is the point where the quadratic and absolute value functions are joining and c is a scaling factor. The original definition of the Huber loss uses $c = \frac{1}{2}$.

Optimization

Backpropagation is an efficient optimization algorithm, where the derivative of the cost function is calculated with respect to each parameter (∂_{ξ_i}) using the chain-rule. The updated parameters in iteration k are obtained by

$$\xi_i^k = \xi_i^{k-1} - \frac{\eta}{M} \sum_{n=1}^{M} \partial_{\xi_i} \{J(\mathcal{N}(\overline{\xi}^{k-1}, T_n), G_n)\}, \tag{2.21}$$

where η is the learning rate, and M is the number of samples used in the update. Higher learning rates yield unstable convergence trapping in local minima, while small values will slow the training down. Depending on M, the below listed optimizations are distinguished.

1. In *gradient descent* each update step uses all samples – M equals the size of the training set. This technique may get trapped in local minima, needs large memory and converges slowly.

2. In *stochastic gradient descent* a random single sample is chosen at a time, $M = 1$, leading to frequent updates and faster convergence. It is computationally less demanding, but is sensitive to outliers and the loss function converges with high variance.

3. The error may be calculated on so-called *batch*es. A small set of sample images (usually $M = 4..64$) are passed through the network, and the loss function is averaged over this set. The ideal batch size uses moderate memory, updates frequently, and yields a relatively smooth convergence.

These definitions give rise to the term *epoch*, the number of optimization steps when the whole training set was once used to update the network.

The *Adam optimizer* [52] is the most popular algorithm in computer vision tasks. It boosts the above gradient descent algorithms with weight-specific adaptation of the learning rate.

$$\xi_i^k = \xi_i^{k-1} - \frac{\eta}{\sqrt{\hat{v}^k} + \epsilon} \hat{m}^k, \tag{2.22}$$

where $\hat{m}$ and $\hat{v}$ are the estimated mean and momentum of the gradient within an exponential moving window. The original paper suggests default settings for the exponential decay and learning rate parameters, which are widely used ever since.

Fully connected network

A conventional design is the *fully connected network*, when all the outputs of a layer are connected to all the neurons of the upcoming layer (Fig. 2.10).

This design leads to an extremely high number of variables, especially in the case of deep networks. Too many tuning parameters compared to the size of the learning dictionary may lead to *overfitting* of the network. It means that the network learns the random noise and nonspecific details of the training set, and is frequently incorrect on unpresented samples, as the learned concepts do not apply to them. This is usually tested with a smaller subset of the learning dictionary, which is not used for error backpropagation. This way the training can be stopped after a given iteration number, avoiding overfitting. Finding an appropriate network structure for

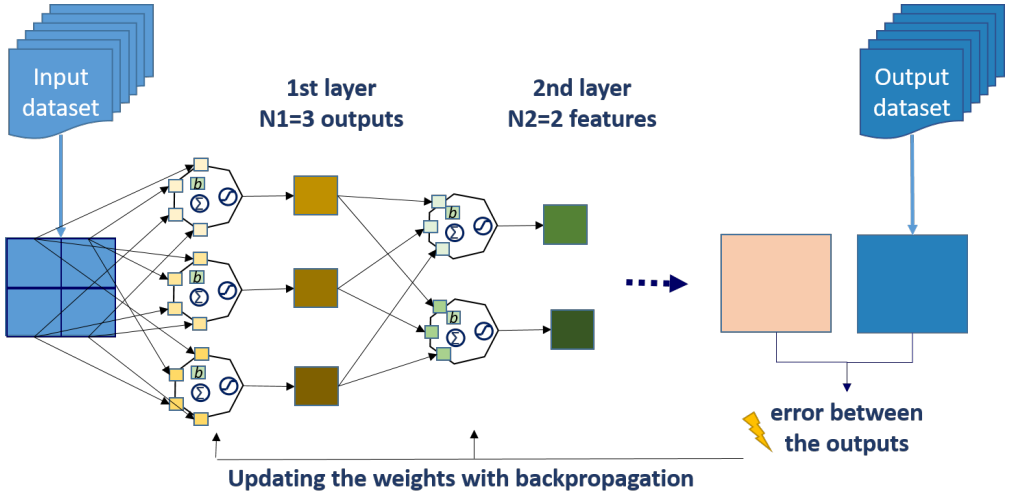

Figure 2.10: In the fully connected neural network all outputs of a layer are inputs to all neurons in the subsequent layer. In this example the input has only 4 pixels, but the first layer already contains 15 tuning parameters, with only 3 outputs, showing the intensive memory requirement of the network design. The final output can be a single or multiple values.

the given learning task - which is complex enough but does not lead to overfitting - is therefore essential.

Convolutional neural network

CNN is a popular neural network structure containing convolutional layers along with other optional layers. The units within a convolutional layer are organized in such a way that the multiplication of input pixels with their corresponding weights implements a convolution process followed by a non-linear activation operator, passing a series of filtered images (*features*) to the upcoming layer (Fig. 2.11). The output of the convolutional layers can be directly an image, or can be reduced to a classification decision using a subsequent fully connected network. This kind of arrangement allows for good feature extraction with a relatively small set of weights, as the extraction of one feature requires only as many weights, as large the kernel size is. An example for a feature in the first layer can be an edge map of the input image. This structure is favored in image processing tasks.

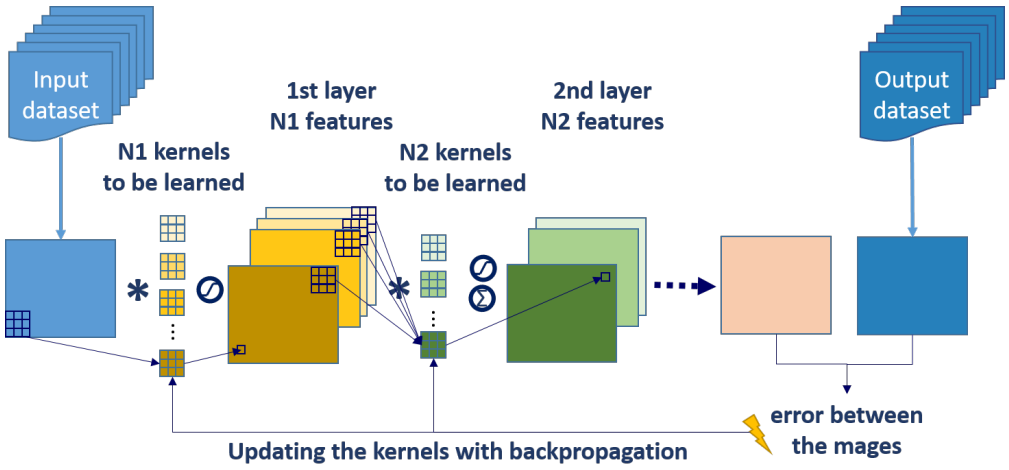

Figure 2.11: The training of the CNN requires a large learning dictionary with known input-output pairs. The weights are represented as convolutional kernels, giving multiple (N1, N2, etc.) outputs (features) in each layer which are passed to a non-linear function. These outputs are the inputs of the next layer. The final output is compared to the expected output, and the error between them is used to update the weights of the network.

Deep learning in the SR problem

Many different approaches have been investigated for the enhancement of training in terms of quality and speed in the SR task. An overview of such techniques is given in [53], and was used as a basis for this summary.

In the pioneering, 2014 work of Dong *et al.* [54] a three layer, rather shallow network, super-resolution CNN (SRCNN) was designed – with convolutional patch extraction, non-linear mapping, and reconstruction (Fig. 2.12 a). Kim *et al.* have shown that the number of layers can increase the performance, so that their very deep SR CNN, VDSR (20 layers with small filters) highly outperforms the shallow ones (Fig. 2.12 b) [55]. A skip connection was also applied between the input and estimated output images in order to learn the residual instead of direct learning of the HR image, giving a more stable convergence. Kim *et al.* have further developed this network using recursive blocks in deeply-recursive convolutional network (DRCN), meaning multiple convolutional layers use the same kernel weights as a way of parameter reduction (Fig. 2.12 c) [56].

However, the previous networks use the upsampled (eg. by bicubic interpolation) LR image as an input, leading to a large number of parameters. A more efficient

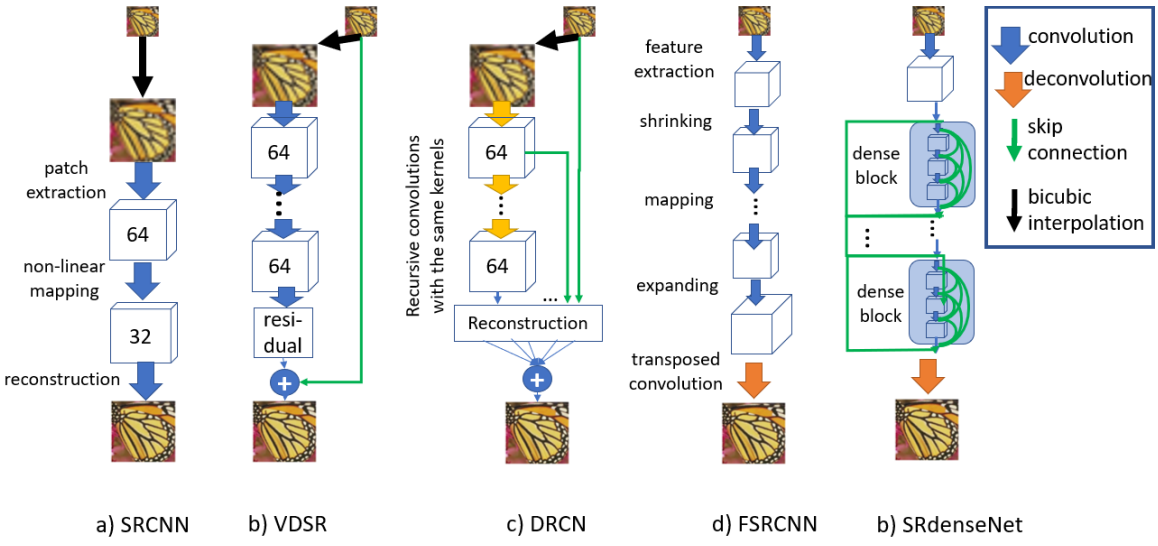

Figure 2.12: State of the art deep learning SR networks. a) The first SR network with 3 convolutional layers. b) Very deep SR network with skip connections and c) recursive convolutional layers. d) Fast SRCNN modifies a) by upscaling the image only in the final layer with transposed deconvolution. e) Uses convolutional blocks with dense skip connections, and upscales the image only in the final layers.

solution, fast SRCNN (FSRCNN) uses transposed convolution as an extension of SRCNN (Fig. 2.12 d) in [57]. In a different solution Shi *et al.* upsample the image in the final layer, re-channeling the multiple features into the HR-image [58, 59]. This structure will be detailed in Chapter 3.

The reduction of parameters using these deconvolution steps, and the efficiency provided by skip connections gave rise to more complex networks. These properties are combined for solving the SISR problem in a densely connected network (Fig. 2.12 e) [60]. The skip connections are concatenating outputs of former blocks, and the output of the dense blocks is fed to a deconvolution layer. Similar skip connections and transposed convolution are used in an interesting structure called U-net, introduced for biomedical image segmentation and artifact removal, where features at different scales are learned efficiently [44, 61]. This structure will be introduced in Chapter 3, the first contribution of this thesis. The Reader is referred to [53] for some further techniques which are out of the scope of this summary, as they were developed in parallel or after our work.

2.3 Tensor algebra

Tensor algebra is used in the proposed methods of Chapter 4. Its basic notations and operations are overviewed in this section. This section builds on [J2], [Au4]. Readers may refer to [62] and [63] for further details.

2.3.1 Tensor operations

Order, fibers, pure tensors

A tensor is a generalization of vectors and matrices, where the *order* of the tensor indicates its dimensionality. A 3D image volume is a third-order tensor $\boldsymbol{X} \in \mathbb{R}^{\mathrm{I}\times\mathrm{J}\times\mathrm{K}}$ from which one dimensional *fibers* can be extracted. Depending on which indices are fixed, there are mode-1 fibers denoted by $\boldsymbol{X}(:, j, k)$ vectors (columns, Fig. 2.13.a), mode-2 fibers denoted by $\boldsymbol{X}(i, :, k)$ vectors (rows, Fig. 2.13.b) and mode-3 fibers denoted by $\boldsymbol{X}(i, j, :)$ vectors (Fig. 2.13.c).

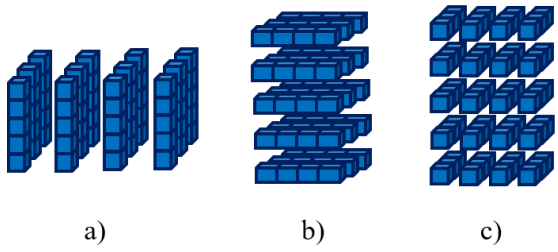

Figure 2.13: Fibers of a third-order tensor - a) mode-1 fibers (columns), b) mode-2 fibers (rows), and c) mode-3 fibers.

The *outer product* (denoted by $\otimes$) between one mode-1, one mode-2 and one mode-3 array forms a rank-1 (also called *pure*) third order tensor, written as

$$\begin{aligned} \boldsymbol{X} =& \mathbf{u} \otimes \mathbf{v} \otimes \mathbf{w}, \text{ where } \boldsymbol{X}(i, j, k) = \mathbf{u}(i)\mathbf{v}(j)\mathbf{w}(k), \\ & \mathbf{u} \in \mathbb{R}^{\mathrm{I}}, \mathbf{v} \in \mathbb{R}^{\mathrm{J}}, \mathbf{w} \in \mathbb{R}^{\mathrm{K}}, \boldsymbol{X} \in \mathbb{R}^{\mathrm{I}\times\mathrm{J}\times\mathrm{K}}. \end{aligned} \tag{2.23}$$

Mode product

Next, the multiplication between a 2D matrix and a 3D tensor referred to as the *mode product* is defined. This multiplication can be performed along all three dimensions, and in each case the mode-n fibers of the tensor are extracted and are

pre-multiplied by the matrix one-by-one. The mode-n products ($n \in \{1,2,3\}$) of $\boldsymbol{X} \in \mathbb{R}^{\mathrm{I}\times\mathrm{J}\times\mathrm{K}}$ with $P_1 \in \mathbb{R}^{\mathrm{I}^*\times\mathrm{I}}, P_2 \in \mathbb{R}^{\mathrm{J}^*\times\mathrm{J}}, P_3 \in \mathbb{R}^{\mathrm{K}^*\times\mathrm{K}}$ are denoted by $\times_n$, and are defined as

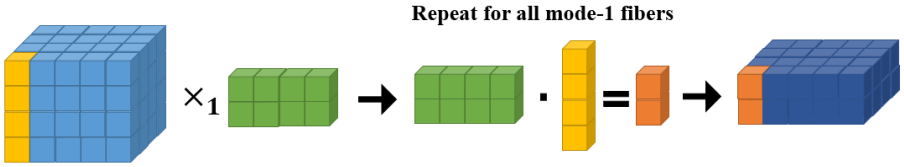

Figure 2.14: Illustration of the mode-1 product. The mode-1 fibers of the 3D tensor are extracted and pre-multiplied by the 2D matrix.

$$\begin{aligned}
\boldsymbol{X} \times_1 P_1 &= \boldsymbol{X_1} \in \mathbb{R}^{\mathrm{I}^*\times\mathrm{J}\times\mathrm{K}}, \text{where } \boldsymbol{X_1}(:,j,k) = P_1\boldsymbol{X}(:,j,k) \\
\boldsymbol{X} \times_2 P_2 &= \boldsymbol{X_2} \in \mathbb{R}^{\mathrm{I}\times\mathrm{J}^*\times\mathrm{K}}, \text{where } \boldsymbol{X_2}(i,:,k) = P_2\boldsymbol{X}(i,:,k) \\
\boldsymbol{X} \times_3 P_3 &= \boldsymbol{X_3} \in \mathbb{R}^{\mathrm{I}\times\mathrm{J}\times\mathrm{K}^*}, \text{where } \boldsymbol{X_3}(i,j,:) = P_3\boldsymbol{X}(i,j,:)
\end{aligned} \tag{2.24}$$

where $\mathrm{I}^*, \mathrm{J}^*, \mathrm{K}^*$ are arbitrary integer numbers. In Fig. 2.14, the principle of the mode-1 product, $\boldsymbol{X} \times_1 P_1 = \boldsymbol{X_1}$ is illustrated, where the columns of the tensor are pre-multiplied by P_1, leading to a shrinkage along the first dimension.

Matricization

Finally, the *matricization* or *unfolding* of the tensor $\boldsymbol{X} \in \mathbb{R}^{\mathrm{I}\times\mathrm{J}\times\mathrm{K}}$ from 3D to 2D is defined. Note that this matricization can be realized again along the three dimensions. For a mode-n matricization the mode-n fibers are extracted and form the columns of $\boldsymbol{X}^{(n)}$ in lexicographical order expressed as

$$\begin{aligned}
\boldsymbol{X}^{(1)} &= [\boldsymbol{X}(:,1,1), \boldsymbol{X}(:,2,1), \ldots \boldsymbol{X}(:,\mathrm{J},1), \boldsymbol{X}(:,1,2) \ldots, \boldsymbol{X}(:,\mathrm{J},\mathrm{K})] \\
\boldsymbol{X}^{(2)} &= [\boldsymbol{X}(1,:,1), \boldsymbol{X}(2,:,1), \ldots, \boldsymbol{X}(\mathrm{I},:,1), \boldsymbol{X}(1,:,2) \ldots, \boldsymbol{X}(\mathrm{I},:,\mathrm{K})] \\
\boldsymbol{X}^{(3)} &= [\boldsymbol{X}(1,1,:), \boldsymbol{X}(2,1,:), \ldots, \boldsymbol{X}(\mathrm{I},1,:), \boldsymbol{X}(1,2,:) \ldots, \boldsymbol{X}(\mathrm{I},\mathrm{J},:)] \\
\boldsymbol{X}^{(1)} &\in \mathbb{R}^{\mathrm{I}\times\mathrm{JK}}, \boldsymbol{X}^{(2)} \in \mathbb{R}^{\mathrm{J}\times\mathrm{IK}}, \boldsymbol{X}^{(3)} \in \mathbb{R}^{\mathrm{K}\times\mathrm{IJ}}.
\end{aligned} \tag{2.25}$$

The mode-1 matricization is illustrated in Fig. 2.15.

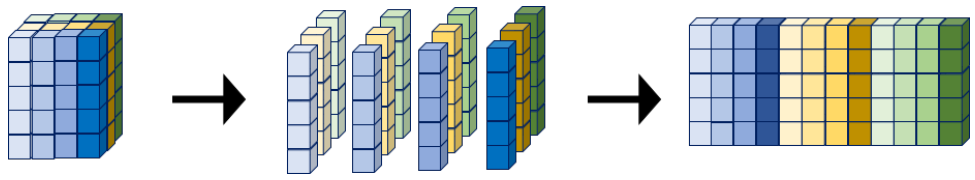

Figure 2.15: Mode-1 matricization of a third-order tensor. The extracted mode-1 fibers are put in lexicographical order to form the 2D matrix.

Decomposition

Matrix and tensor *decomposition* (also called factorization) are operations that help to decompose a matrix or tensor into constituent elements, which offer easier implementations of complex algorithms. Solving a system of linear equations, or calculating the inverse are problems calling for such decompositions.

A group of tensor decomposition are different generalizations of the matrix singular value decomposition (SVD), which is

$$A = U\Sigma V^* \tag{2.26}$$

where U and V are unitary matrices, Σ is a rectangular diagonal matrix, and (V^*) is the complex conjugate of V. The diagonal elements of Σ are the singular values of the matrix, and the number of non-zero entries is the rank of the matrix. Following the different generalizations it is shown in the upcoming sections, that higher order tensors have different definitions of ranks depending on the factorization. Kolda and Bader gave a comprehensive summary of these generalized decomposition techniques in [63], hereafter two of them will be introduced.

2.3.2 Canonical polyadic decomposition

The smallest number of rank-1 tensors that can sum up to form tensor $\boldsymbol{X}$ is called the *tensor rank* of $\boldsymbol{X}$, denoted by F. The resulting factorization of $\boldsymbol{X}$ is called the *canonical polyadic decomposition* (CPD) of $\boldsymbol{X}$ expressed as

$$\begin{aligned} \boldsymbol{X} &= \sum_{f=1}^{F} U_1(:,f) \otimes U_2(:,f) \otimes U_3(:,f), \\ \text{where } \boldsymbol{X}(i,j,k) &= \sum_{f=1}^{F} U_1(i,f) U_2(j,f) U_3(k,f). \end{aligned} \tag{2.27}$$

$\overline{U} = \{U_1, U_2, U_3\}$ is a set of three 2D matrices, $\{U_1 \in \mathbb{R}^{\mathrm{I}\times F}, U_2 \in \mathbb{R}^{\mathrm{J}\times F}, U_3 \in \mathbb{R}^{\mathrm{K}\times F}\}$, known as the decomposition of the tensor $\boldsymbol{X}$. For illustration, the reader may refer to Fig. 2.16. In the following, the operation in (2.27) will be denoted by

$$[\![U_1, U_2, U_3]\!] = \sum_{f=1}^{F} U_1(:, f) \otimes U_2(:, f) \otimes U_3(:, f). \tag{2.28}$$

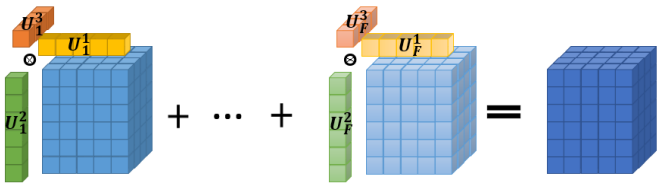

Figure 2.16: Illustration of the CPD. F is the number of outer products formed by mode-1 ($U_i^1 := U_1(:, i)$), mode-2 ($U_i^2 := U_2(:, i)$) and mode-3 ($U_i^3 := U_3(:, i)$) fibers summing up to a tensor with rank F.

Using the factorization of $\boldsymbol{X}$ in (2.27) and (2.28) the mode-n products (2.24) can also be written as

$$\boldsymbol{X} \times_1 P_1 \times_2 P_2 \times_3 P_3 = [\![P_1U_1, P_2U_2, P_3U_3]\!]. \tag{2.29}$$

The matricization can also be realized using the CPD $\overline{U}$ of $\boldsymbol{X}$. For this the column-wise *Khatri-Rao product* is necessary (denoted by $A \odot B = C$, where $A \in \mathbb{R}^{\mathrm{I}\times F}, B \in \mathbb{R}^{\mathrm{J}\times F}, C \in \mathbb{R}^{\mathrm{IJ}\times F}$). It operates on matrices having the same number of columns, and calculates their column-wise *Kronecker-product* (vectorized outer product). Using this notation the matricization (2.25) can be written as

$$\begin{aligned} \boldsymbol{X}^{(1)} &= U_1(U_3 \odot U_2)^T \\ \boldsymbol{X}^{(2)} &= U_2(U_3 \odot U_1)^T \\ \boldsymbol{X}^{(3)} &= U_3(U_2 \odot U_1)^T. \end{aligned} \tag{2.30}$$

This matricization and (2.29) allows the construction of efficient algorithms for finding the corresponding CPD. Usually an *alternating least square* (ALS) algorithm is

used, as proposed in [64]. The subproblems are

$$\begin{aligned}
&\min_{U_1} ||\boldsymbol{X}^{(1)} - U_1(U_3 \odot U_2)^T||_{\mathrm{F}} \\
&\min_{U_2} ||\boldsymbol{X}^{(2)} - U_2(U_3 \odot U_1)^T||_{\mathrm{F}} \\
&\min_{U_3} ||\boldsymbol{X}^{(3)} - U_3(U_2 \odot U_1)^T||_{\mathrm{F}},
\end{aligned} \tag{2.31}$$

using the Frobenius-norm $\|\cdot\|_F$ defined for a third order tensor as

$$||\boldsymbol{X}||_{\mathrm{F}} = \left(\sum_{i=1}^{\mathrm{I}} \sum_{j=1}^{\mathrm{J}} \sum_{k=1}^{\mathrm{K}} \left(\boldsymbol{X}\left(i,j,k\right)\right)^2 \right)^{1/2} \tag{2.32}$$

A solution for (2.31) is given by

$$\begin{aligned}
U_1 &= \boldsymbol{X}^{(1)}[(U_3 \odot U_2)^T]^\dagger \\
U_2 &= \boldsymbol{X}^{(2)}[(U_3 \odot U_1)^T]^\dagger \\
U_3 &= \boldsymbol{X}^{(3)}[(U_2 \odot U_1)^T]^\dagger,
\end{aligned} \tag{2.33}$$

where $(^\dagger)$ is a pseudoinverse. In the algorithm only one dimension is updated at a time, the others are fixed, until the solution converges.

An important property of the CPD is that this decomposition is essentially unique (allowing permutations within U_1, U_2, U_3). Thus $\overline{U}$ can be identified almost surely if its tensor rank F is smaller than an upper bound. Chiantini *et al.* [65] proved that if $\mathrm{I} \geq \mathrm{J} \geq \mathrm{K}$, with $F \leq 2^{\lfloor log_2 \mathrm{J} \rfloor + \lfloor log_2 \mathrm{K} \rfloor - 2}$, the CPD of the rank-$F$ tensor $\boldsymbol{X} \in \mathbb{R}^{\mathrm{I} \times \mathrm{J} \times \mathrm{K}}$ is essentially unique. This condition allows identifiability of the CPD even for tensors with high rank. For example, a typical CT volume with $260 \times 260 \times 300$ pixels can be decomposed uniquely even if the tensor rank of the image is as high as $2^{14} = 16384$.

However, a main difference between the matrix rank and the above tensor rank F is that for the calculation of the tensor rank there is no straightforward algorithm, as this task is NP-hard. This explains why a relaxed form of the tensor rank, (2.11) is used in the LRTV method. Usually multiple decompositions are tried with fixed R number of components to approximate $\boldsymbol{X}$

$$\boldsymbol{X} \approx \hat{\boldsymbol{X}} = [\![\hat{U}^1, \hat{U}^2, \hat{U}^3]\!] = \sum_{i=1}^{R} \hat{U}^1(:,i) \otimes \hat{U}^2(:,i) \otimes \hat{U}^3(:,i). \tag{2.34}$$

If $R < F$, a denoising or data compression might be realized.

The original algorithm [64] was used in psychometrics, analyzing multiple similarity, dissimilarity matrices. It was also used in the modeling of fluorescence excitation-emission data [66], in the analysis of fMRI data [67] and time-varying electroencephalogram data [68].

With some modification the algorithm was used for image compression and classification [69], data fusion in multi- and hyperspectral imaging [62]. This fusion technique combines a multispectral image (with high spatial and low spectral resolutions) and a hyperspectral image (with low spatial and high spectral resolution) to obtain an SR image (with high spatial and high spectral resolutions). One advantage of the tensor-based method of [62] is that it does not need to unfold the image of interest into a 2D matrix [29], or take only slices of the volume [J1]. As a consequence, this method avoids any loss of information about the locality of the image pixels and does not require to introduce spatial regularization (such as the TV of the image). Our proposed algorithm for SISR uses a similar method in Section 4.2 and 4.3.

2.3.3 Tucker decomposition

The *n-rank* of a tensor is the number of its independent mode-n fibers. It can be obtained as the rank of the mode-n matricized tensor. For a third-order tensor three such values, {1-rank, 2-rank, 3-rank} can be defined. These n-ranks should not be confused with the rank F of the tensor, defined by the CPD.

The *Tucker decomposition* TD (also called higher order SVD, or multi-linear SVD) is the multidimensional generalization of the 2D SVD, written as

$$\boldsymbol{X} = \boldsymbol{\Sigma} \times_1 V_1 \times_2 V_2 \times_3 V_3, \tag{2.35}$$

where $\boldsymbol{\Sigma} \in \mathbb{R}^{R_1 \times R_2 \times R_3}$ is a *core tensor*, and $V_1 \in \mathbb{R}^{\mathrm{I} \times R_1}, V_2 \in \mathbb{R}^{\mathrm{J} \times R_2}, V_3 \in \mathbb{R}^{\mathrm{K} \times R_3}$ are the (usually orthonormal) bases of the subspaces spanned by the mode-n fibers [63]. If R_n (in particular R_1, R_2, R_3) equal the n-ranks of $\boldsymbol{X}$, the decomposition is exact, but for lower numbers and in the presence of noise it becomes inexact, and gives only an approximation $\tilde{\boldsymbol{X}}$. The CPD can be viewed as a special case of the

TD, with $R_1 = R_2 = R_3$.

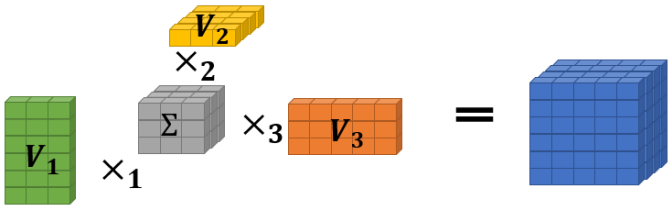

Figure 2.17: Illustration of the exact Tucker decomposition, a generalization of the 2D SVD. A core tensor $\boldsymbol{\Sigma}$ *is multiplied along the corresponding modes by base matrices* V_1, V_2, V_3.

As opposed to the 2D SVD, $\boldsymbol{\Sigma}$ is not diagonal, but its elements show the level of interaction between the different modes. Due to orthonormality, it can be obtained as

$$\boldsymbol{\Sigma} = \boldsymbol{X} \times_1 V_1^T \times_2 V_2^T \times_3 V_3^T, \tag{2.36}$$

The shorthand of TD is

$$\boldsymbol{X} = [\![\Sigma; V_1, V_2, V_3]\!]. \tag{2.37}$$

Singular values similar to the 2D case (SV_n for $n = 1, 2, 3$) can be defined, showing the relevance of the individual base vectors. These are calculated as the Frobenius norms (2.32) of 2D slices taken from the core tensor $\boldsymbol{\Sigma}$, along mode-n, fixing the index of the component in question as

$$\begin{aligned} SV_1(i) &= \|\boldsymbol{\Sigma}(\boldsymbol{i}, :, :)\|_F \\ SV_2(j) &= \|\boldsymbol{\Sigma}(:, \boldsymbol{j}, :)\|_F \\ SV_3(k) &= \|\boldsymbol{\Sigma}(:, :, \boldsymbol{k})\|_F . \end{aligned} \tag{2.38}$$

When $R_n <$ n-rank$(\boldsymbol{X})$, it is called the truncated TD. While for 2D matrices the truncation of the SVD gives an optimal approximation of the matrix in the least square sense, it is not true for higher order tensors, but gives a good starting point for further ALS algorithms. Furthermore, TD is not unique, giving flexibility in the composition of $\boldsymbol{\Sigma}$.

The algorithm was first introduced and further improved by Tucker and Levin [70]. In some cases the decomposed form requires significantly less memory than the original tensor does [71]. The decomposition was also used in facial recognition in [72], for image fusion to overcome the problem of limited depth-of-field

camera pictures [73], for the resolution enhancement of color images [74]. The truncated version was applied for the denoising of MRI scans [75]. TD was used in the hyperspectral-multispectral super-resolution problem with success [76]. The proposed SISR algorithm in Section 4.4 also uses TD.

Chapter 3

Proposed deep learning - based image enhancement

The resolution of dental CBCT images is limited by detector geometry, sensitivity, patient movement, the reconstruction technique and the need to minimize the radiation dose. Recently, the use of CNN architectures has shown promise as a resolution enhancement method. In the current work, two CNN architectures – a subpixel network and the so called U-net – have been considered for the resolution enhancement of 2D CBCT image slices of *ex vivo* teeth. To do so, a training set of 5680 cross-sectional slices of 13 teeth and a test set of 1824 slices of 4 structurally different teeth were used. Two existing reconstruction-based super-resolution methods using ℓ_2-norm and TV regularization were used for comparison. The results were evaluated with different metrics (peak signal-to-noise ratio, structure similarity index, and other objective measures estimating human perception) and subsequent volume-segmentation-based analysis. In the evaluation, µCT images were used as ground truth. The results suggest the superiority of the proposed CNN-based approaches over reconstruction-based methods in the case of CBCT images, allowing better detection of medically salient features such as the size, shape, or curvature of the root canal.

3.1 Experimental setup

This section introduces the CT images, metrics, PSF-estimation method used in the contributions of this thesis.

3.1.1 Dataset

As the methods presented in this thesis were tested on the same dataset, the images will be described in this section.

Images of 17 intact freshly extracted teeth (incisors, canines, premolars, and molars for structural diversity) were acquired. These teeth were donated anonymously for research and had been extracted for reasons unrelated to the current study. A Carestream 81003D limited CBCT system, currently used in clinics, was used for the LR image acquisition, and a Quantum FX µCT system from Perkin Elmer for the HR images. Imaging was performed at Life Imaging Facility of Paris Descartes University (Plateforme Imageries du Vivant - PIV) on µCT Platform site (EA2496, Montrouge, France). The resolution of the CBCT machine was 1 LP/mm at 50% MTF (this term is explained in Section 2.1.3), defining a linewidth of 500 µm. The reconstructed voxel size was 75 μm^3. For the µCT machine the resolution was 10 LP/mm at 50% MTF (a linewidth of 50 µm), the reconstructed voxel size was 40 μm^3.

The acquired CBCT images were automatically registered onto the µCT volume with the 3D Slicer tool [77–79], using linear interpolation in the rescaling step. Note that in addition to being geometrically aligned, both sets of images had a common voxel size of 40 μm^3 after the registration process. The axial cross-sectional slices were saved as single images for both types of volumes. The reason for transforming the CBCT images to the pixel resolution of the µCT images (rather than the other way round) was to avoid degradation of the intrinsic resolution of the µCT images and thereby reducing the training sample number.

The noise and the reconstruction errors in the background of the images are structurally different on the two modalities. This difference is investigated in Fig. 3.1 where on a log-scale the noise in the background is clearly visible, both on the

CBCT and on the µCT images. On the mean histograms of the training sets it can be seen that the pixels of the background and those of the foreground are easily separable with global thresholds (dashed lines). The result of this thresholding can be seen on the example images as masks of the tooth. After thresholding, the pixel intensities were normalized between 0 and 1. It was qualitatively and quantitatively investigated how the deep learning SR algorithms handle this difference in noise patterns, and how do they perform after background removal.

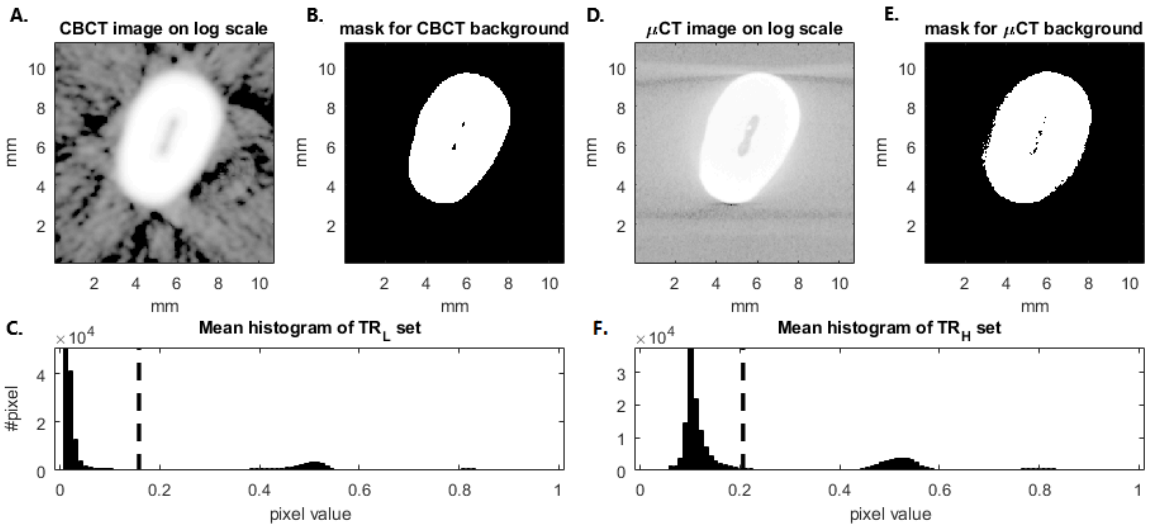

Figure 3.1: Background artifacts. A.: background artifacts on the CBCT image on log scale B.: mask of the tooth from A. C.: mean histogram of CBCT images from $\overline{TR}_L$, with the global threshold used for background removal (dashed line). The same threshold was used for masking on B. D.: background artifacts on the µCT image on log scale E.: mask of the tooth from D. F.: mean histogram of µCT images from $\overline{TR}_H$, with the global threshold used for background removal (dashed line). The same threshold was used for masking on E.

,

13 teeth were selected for the training sets, and four other teeth (an incisor, a premolar and two molars) provided the test sets. The training set of low-resolution CBCT and high-resolution µCT image slices are denoted by $\overline{TR}_L$ and $\overline{TR}_H$, and the corresponding test sets by $\overline{TE}_L$ and $\overline{TE}_H$, and the corresponding sets of the 3D volumes by $\overline{\boldsymbol{TR}}_{\boldsymbol{L}}$, $\overline{\boldsymbol{TR}}_{\boldsymbol{H}}$, $\overline{\boldsymbol{TE}}_{\boldsymbol{L}}$, $\overline{\boldsymbol{TE}}_{\boldsymbol{H}}$.

3.1.2 PSF estimation for the model-based algorithms

The blurring effect is generally modelled as a convolution with a spatially invariant PSF. In practice the PSF is unknown, so it needs to be measured or estimated. Its measurement for experimental data is a very difficult task, therefore it is estimated from data in many existing works (see [32] for an example). The SISR degradation model (2.4) assumes that the CBCT images are low-pass filtered versions of the ideal µCT images, where the blurring kernel is Gaussian and its standard deviation was estimated from the observed data. Employing direct inverse filtering on each sample image, the µCT volume was divided frequency-wise by the CBCT volume in the Fourier-domain, to obtain the PSF (the constant λ was used to avoid division by zero). The high-frequency band was suppressed by a 3D Hamming-window before computing the inverse Fourier-transform. The training sets were used to estimate an averaged PSF, in order to reduce noise. The 3D PSF was thus obtained as:

$$\begin{aligned} \mathcal{F}(\mathrm{PSF}) &= \frac{1}{|\overline{\boldsymbol{TR}}_{\boldsymbol{L}}|} \sum_{k \in \overline{\boldsymbol{TR}}_{\boldsymbol{L}}} \mathcal{F}(\overline{\boldsymbol{TR}}_{\boldsymbol{L}}\{k\}) \oslash [\mathcal{F}(\overline{\boldsymbol{TR}}_{\boldsymbol{H}}\{k\}) + \lambda \cdot J], \\ \mathrm{PSF} &= \mathcal{F}^{-1}(\mathrm{Hamming}(\mathrm{PSF})). \end{aligned} \tag{3.1}$$

where $\mathcal{F}$ denotes the Fourier transform operator, $|\overline{\boldsymbol{TR}}_{\boldsymbol{L}}|$ is the cardinality of the set, k is the training image index, $\oslash$ is the element-wise Hadamard division, λ is a small positive real number, and J is a matrix of ones having the same size as the images. The division here is to be considered element-wise. A Hanning-window was applied to the estimated PSF, suppressing high-frequency noise due to edge effects.

Finally a 3D Gaussian function was fitted to the averaged PSF to estimate the standard deviations $\overline{\sigma}$ ($\{\sigma_1, \sigma_2, \sigma_3\}$ for the three dimensions).

3.1.3 Metrics

As LP/mm resolution of the images can be accurately calculated only on the standard phantom (2.1.3), the enhancement of the biological images has to be measured with other metrics.

The first one measures the *peak signal-to-noise ratio* (PSNR) between the input $\boldsymbol{A}$ and reference $\boldsymbol{R}$ (µCT, whose values have been normalized between 0 and 1) image volumes. It is calculated by dividing the square of the dynamic range (difference between the maximum and minimum values of the input image, $max_{\boldsymbol{A}} - min_{\boldsymbol{A}}$) with the *mean square error* (MSE) between $\boldsymbol{A}$ and $\boldsymbol{R}$, expressed in dB as

$$\mathrm{PSNR} = 10 log_{10} \frac{(max_{\boldsymbol{A}} - min_{\boldsymbol{A}})^2}{\mathrm{MSE}_{\boldsymbol{A/R}}}, \tag{3.2}$$

where

$$\mathrm{MSE}_{\boldsymbol{A/R}} = \frac{1}{\mathrm{I} \cdot \mathrm{J} \cdot \mathrm{K}} \sum_{i=1}^{\mathrm{I}} \sum_{j=1}^{\mathrm{J}} \sum_{k=1}^{\mathrm{K}} (\boldsymbol{A}(i,j,k) - \boldsymbol{R}(i,j,k))^2. \tag{3.3}$$

The *structural similarity index* (SSI) was designed to better reflect subjective evaluation [80]. It combines the luminance ($l(i,j,k)$), contrast ($c(i,j,k)$), and structural measures ($s(i,j,k)$), and can be calculated for each pixel as the SSI map (SSIM) as

$$\mathrm{SSIM}(\boldsymbol{A}, \boldsymbol{R}) = [l(\boldsymbol{A}, \boldsymbol{R})] \cdot [c(\boldsymbol{A}, \boldsymbol{R})] \cdot [s(\boldsymbol{A}, \boldsymbol{R})], \tag{3.4}$$

with

$$\begin{aligned} l(\boldsymbol{A}, \boldsymbol{R})] &= \frac{2\mu_{\boldsymbol{A}}\mu_{\boldsymbol{R}} + C_1}{\mu_{\boldsymbol{A}}^2 + \mu_{\boldsymbol{R}}^2 + C_1}, \\ c(\boldsymbol{A}, \boldsymbol{R})] &= \frac{2\sigma_{\boldsymbol{A}}\sigma_{\boldsymbol{R}} + C_2}{\sigma_{\boldsymbol{A}}^2 + \sigma_{\boldsymbol{R}}^2 + C_2}, \\ s(\boldsymbol{A}, \boldsymbol{R})] &= \frac{2\sigma_{\boldsymbol{AR}} + C_3}{\sigma_{\boldsymbol{A}}\sigma_{\boldsymbol{R}} + C_3}, \end{aligned} \tag{3.5}$$

where $\mu_{\boldsymbol{A}}$, $\mu_{\boldsymbol{R}}$, $\sigma_{\boldsymbol{A}}$, $\sigma_{\boldsymbol{R}}$, $\sigma_{\boldsymbol{AR}}$ are local means, standard deviations and cross-covariance taken in small neighborhoods of each pixel. C_1, C_2, and C_3 are regularization values to avoid division by zero. For the whole image an average of the single values can be taken, denoted by SSI.

The *information fidelity criterion* (IFC) quantifies the mutual information between two images. The images are first transformed into the wavelet domain, and a Gaussian scale mixture model is fitted to the coefficients. The conditional mutual information is calculated between these models. The resulting measure is correlating with the human perceptual quality without parameters [81].

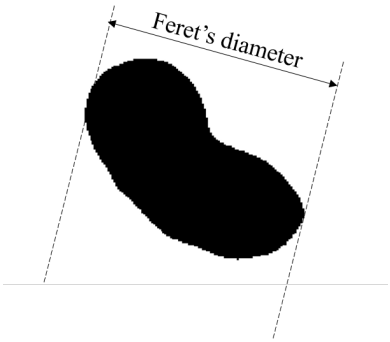

Figure 3.2: The Feret's diameter is the longest distance between two parallel straight lines that are tangent to the shape.

The *noise quality measure* (NQM) analyzes $\boldsymbol{A}$ and $\boldsymbol{R}$ using a contrast pyramid in the spatial and frequency domains [82]. It quantifies spatially varying non-linear visual effects as variation in contrast sensitivity and local luminance, contrast interaction between spatial frequencies, and contrast masking effects.

The enhanced images were also compared as 3D volumes for a more application-oriented evaluation. The canal root was segmented from the 3D volume $(\boldsymbol{A}_{canal}, \boldsymbol{R}_{canal})$ using a dedicated adaptive local thresholding described in [83]. For each root, the canal area and the *Feret's diameter* were estimated for all the radicular axial reconstructions, as suggested in [83]. The Feret's diameter defines the longest distance between two parallel straight lines that are tangent to the shape, as shown in Fig. 3.2. The comparison of Feret's diameter on corresponding slices is evaluated using the mean of differences. It shows whether there is a systematic error or bias between the two images. The segmented volumes were also measured, showing the absolute differences with the ground-truth µCT images in percentages, and using the Dice coefficient [84]

$$\text{Dice}(\boldsymbol{A}_{canal}, \boldsymbol{R}_{canal}) = 2 \cdot \frac{|\boldsymbol{A}_{canal} \cap \boldsymbol{R}_{canal}|}{|\boldsymbol{A}_{canal}| + |\boldsymbol{R}_{canal}|}, \tag{3.6}$$

where $|\boldsymbol{A}_{canal} \cap \boldsymbol{R}_{canal}|$ is the number of common canal-pixels of segmented $\boldsymbol{A}$ and $\boldsymbol{R}$, and $|\boldsymbol{A}_{canal}| + |\boldsymbol{R}_{canal}|$ are the number of all canal-pixels in $\boldsymbol{A}$ and $\boldsymbol{R}$.

Finally MeVisLab [85] was used for visualizing the segmentation results.

3.2 Realizations of the CNN

To investigate the potentials of deep learning in dental CT image enhancement, two architectures of CNNs were created. In the discussion that follows, the organization of the layers for each of the two architectures will be first presented, followed by a description of the error metric used to train the networks. The term features will refer to channels of the CNN along the usual definition of its processing pipeline which act as implicit features in the reconstruction process.

3.2.1 U-net

One of the investigated architectures was inspired by the U-net architecture [61] which is commonly used for domain-to-domain transformation, especially in medical imaging. The number of pixels was the same in the $\overline{TR}_L$–$\overline{TR}_H$ and $\overline{TE}_L$–$\overline{TE}_H$ image pairs after the registration step of Section 3.1.1, but the whole dataset had to be scaled to the same dimensions (400×400 pixels) as the network uses a universal weight-structure for all the inputs.

The implemented structure contains four successive downsampling layers on the original input image, continued by four upsampling steps which were implemented by transposed convolutions. This arrangement grants feature extraction using two-two convolutional layers at different scales. At each size-level lateral connections concatenating the downsampled image features to the upsampled ones were also made, keeping feature information of all scales, combining local and global features.

Our implementation altered the original U-net by completing the convolutional layers (Section 2.2.3) with circular boundary conditions, *batch normalization* (fixing the input mean and variance between convolutions [86]) and LReLU (Section 2.2.3), with $\alpha = 10^{-3}$ as defined in (2.16). The number of convolutions – different features – were half of those published in the original paper in order to better fit our small dataset: 32,64,128,256 in the downsampling layers, 256,128,64,32 in the upsampling layers, and two convolutions with 512 features in the lowest resolution.

It has been shown in various problems that the application of smaller kernel sizes can result in a lower number of parameters and higher accuracy [87]. Therefore, the

size of all the kernels employed here was 3×3.

A detailed block diagram of this architecture can be seen in Fig. 3.3.

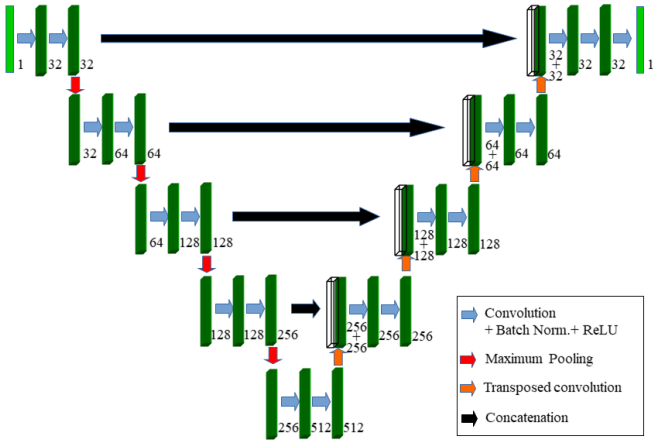

Figure 3.3: Depiction of the U-net structure as a domain-to-domain transformation converting the input image with size 400x400x1 to an image of similar shape but different features. As it can be seen on the figure this structure is good to process local and global features together. The neurons in the deeper layers have larger and larger receptive fields. The numbers in the bottom right corner of the layers indicate the number of features stored. Image was modified from [61].

3.2.2 Subpixel network

The second architecture evaluated for image enhancement was motivated by the subpixel networks implemented by Shi [58], where deconvolution is realized as a tiling operator, instead of transposed convolutions [59]. Here all the inputs of $\overline{TR}_L$ and $\overline{TE}_L$ were scaled to 200 $\times$ 200 pixels, while the expected outputs of $\overline{TR}_H$ and $\overline{TE}_H$ to 400 $\times$ 400 pixels.

A commonly used six layer CNN structure was implemented, as opposed to the three layers of the original implementation. It contained convolution, ReLU, and batch normalization in each layer, with 16, 32, 32, 64, 64, 4 features respectively. The last layer with four features is needed for the depth-to-space operation to give space to the higher resolution on a higher number of pixels (by a factor of two compared to the image size of the original input). The *retiling* operation that rearranges the elements of an $\mathrm{I} \times \mathrm{J} \times \mathrm{K}r^2$ tensor $\boldsymbol{A}$ to a tensor with a shape of $(r\mathrm{I} \times r\mathrm{J} \times \mathrm{K})$ – and as such is responsible for the upscaling – can be defined as:

$$\text{retiling}\left\{\boldsymbol{A}\right\}(x,y,z) = \boldsymbol{A}(\lfloor x/r \rfloor, \lfloor y/r \rfloor, \mathrm{K} \cdot r \cdot \mathrm{mod}(y,r) + \mathrm{K} \cdot \mathrm{mod}(x,r) + z), \qquad (3.7)$$

where x, y and z are the width, height and depth indices of the input image, r is the upsampling factor (in this case 2), K is the input depth of the image, $\lfloor \cdot \rfloor$ is the floor function, and mod$(\cdot)$ is the modulo operation. The processing pipeline of this retiling can be seen in Fig. 3.4.

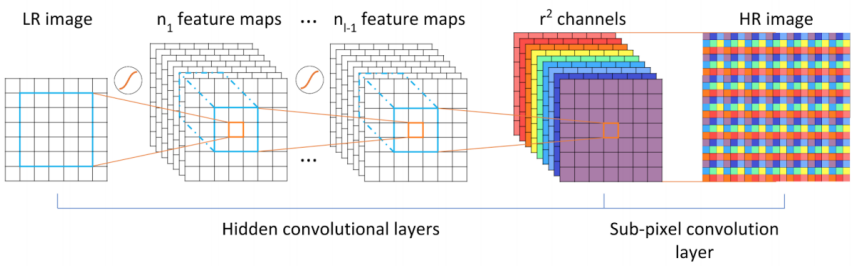

Figure 3.4: Depiction of the retiling (depth-to-space) operation which was tested for enhancing image quality. The image was taken from [59], showing an upsampling factor of three.

3.2.3 Training

The training sets $\overline{TR}_L$ and $\overline{TR}_H$ contained 5680 slices of the 13 teeth, and four other teeth (an incisor, a premolar and two molars) provided 1824 slices for the test sets $\overline{TE}_L$ and $\overline{TE}_H$. In spite of the small number of teeth, the large variability of the slices allowed more precise measurements on a greater set of independent 2D images. The CBCT and µCT images were uniformly normalized using the highest and lowest pixel values found in the training sets $\overline{TR}_L$ and $\overline{TR}_H$ accordingly.

For training the networks on the $\overline{TR}_L$ and $\overline{TR}_H$ sets, the Adam optimizer algorithm (Section 2.2.3) was used with dynamic learning rate initially set to 10^{-4}. The network was trained with randomly initialized weights using the Xavier method as it is described in [88], and there were no significant differences in training depending on the weight and parameter settings. Similarly the initial learning rate of the used Adam optimizer did not have effect on overall reconstruction accuracy of the network. Random batches of 64 images for the subpixel architecture and those of 16 images for the U-net structure were used.

For visualizing the evolution of error between the expected and network outputs, the ℓ_1 loss and the the Huber loss ℓ_{1s} were implemented (as introduced in Section 2.2.3, (2.20) with $c = 1$, $\delta = 1$). For the Adam optimizer the ℓ_{1s} loss was set. Note

that for image normalization in the first layers and also between the layers batch normalization was employed [86]. This method ensures that input data in training batches is transformed to zero mean and unit variance. Examining the ℓ_{1s} loss in (2.20) it can be seen, that those images and regions where a larger variance appeared fall into the $|O - G| \geq 1$ region, penalizing the difference moderately. To the best of our knowledge this method is the most commonly used normalization method for deep learning image applications.

The neural networks were realized using the open-access deep learning framework TensorFlow 1.3.0 [89], running on an NVIDIA GK210GL (Tesla K80 with 12 GB RAM) GPU. The training of a network took two days. In inference mode both trained networks super-resolve a slice in 0.5 s, a whole volume in 2.5 min on average.

The structure of the two networks along with the algorithms and chosen parameters used for the training can be found on GitHub [link to source code].

3.3 Results

Four SISR algorithms – the fast SISR algorithm (Section 2.2.2) with ℓ_2 (SRR-ℓ_2)-norm, and TV (SRR-TV) regularizations, and CNNs with the subpixel and U-net architecture – were applied on the test set, $\overline{TE}_L$ for evaluation of the proposed methods.

3.3.1 Evolution of the loss function

Figure 3.5 shows a comparison of convergences regarding the loss using the ℓ_1 (upper plot) and ℓ_{1s} (lower plot) functions. The network was trained for 20 epochs, the loss converged and did not change significantly after 10 epochs. The reconstruction error of the U-net architecture was much lower using both loss functions, but as it will be discussed later, this result does not agree with the conclusion based on image quality metrics. When applying background removal, the loss values of the networks decrease for both loss functions (Fig. 3.5). Thus, background removal helped both networks to decrease the reconstruction error and to speed up the convergence.

3.3.2 Effect of background noise

The effect of the background noise was investigated qualitatively and quantitatively. An example slice can be seen in Fig. 4.2 for qualitative evaluation. It can be observed, that the SRR methods led to an amplification of this error, and were also causing artifacts on the edges. This latter phenomenon persisted after background removal too. On the other hand, the CNNs – especially the U-net – learned the shape of the background-noise on the µCT image, but estimated a blurred version of its pattern. As the lower row indicates, this problem can also be solved with background removal.

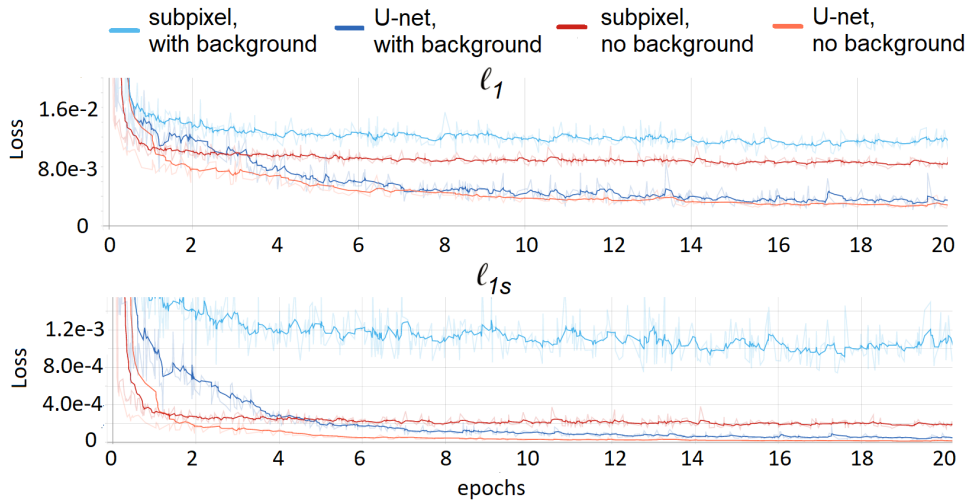

Figure 3.5: Loss of the different networks during training according to the ℓ_1 and ℓ_{1s} metrics. To help visualize the general trends without the short-time randomness of the training algorithm, exponentially smoothed values are shown in dark, and the original values are plotted in semi-transparent colors.

The quantitative effect of the background noise can be seen in Table 3.1. The values were calculated against the ground truth images on the four teeth of the test set. All the measures apart from the IFC showed an improvement after background removal.

The first value to consider is the NQM, as it directly shows the quality of the noise. This value increased significantly with background removal for all the methods. When calculating the MSE, the differences on the relatively large area of the background led to a high error-rate, and thus to a lower PSNR. It also caused a

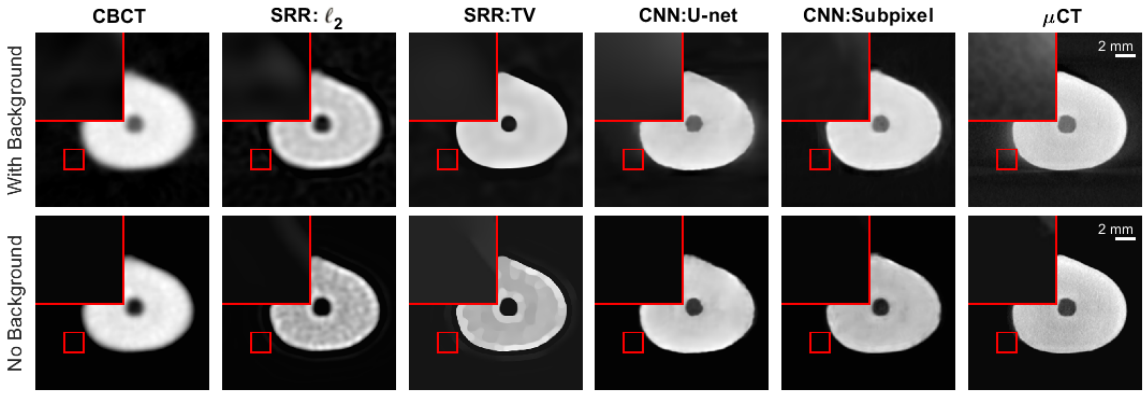

Figure 3.6: Effect of background removal on noise amplification. The columns stand for the four different enhancement methods along with the original CBCT and μCT images. The upper row shows an example slice with intact background, while on the lower row background-removal was carried out. It can be seen that the SRR methods amplify the noise, while the deep learning methods are trying to learn the background-pattern of the μCT image. After background removal this problem no longer holds, only edge-effects of the SRR methods can be observed. The display range was stretched to [0,1].

higher (and different) variance of the compared backgrounds, which effected negatively the SSI values. This effect is less significant on the results with the CNNs, as they learned a similar noise pattern. The decrease of the IFC value following background removal is supposed to be due to the decrease in image variance.

When performing SRR-ℓ_2 and SRR–TV methods not all measures show improvement after background removal. Sometimes background removal seems to negatively affect reconstruction at the edges. Although it results in a visually better image with higher contrast, not all quality metrics can capture this improvement. The deep learning methods, however, do not suffer from this effect, significantly outperforming the traditional methods in every case - even when the contrast is lower than that of the SRR images. It should be noted, that the contrast of the CNN methods is still higher than on the CBCT images.

Table 3.1: Quantitative DL enhancement results

Average Values of PSNR (dB), SSI, IFC and NQM for the Test Set compared to the µCT images. Best Results are Marked in Bold

Metric	background	CBCT	SRR:ℓ_2	SRR:TV	CNN:U-net	CNN:Subpixel
PSNR	yes	22.48	23.62	23.60	23.79	**24.50**
	no	45.56	64.15	64.80	**67.58**	66.60
SSI	yes	0.3801	0.5474	0.5869	0.8045	**0.8182**
	no	0.9145	0.8688	0.8830	0.9304	**0.9346**
IFC	yes	0.3217	0.3348	0.3313	0.5472	**0.5536**
	no	0.2605	0.1908	0.2268	0.4159	**0.4186**
NQM	yes	6.93	7.26	6.85	8.07	**8.64**
	no	9.28	8.02	8.43	9.93	**11.54**

3.3.3 Resolution enhancement on background-removed images

As the validity of background-removal was qualitatively and quantitatively discussed, from hereon only the results obtained with the modified (without background) images will be examined. The values of Table 3.1 confirm the superiority of the proposed deep learning-based methods. The average PSNR increased by 18.59 and 19.24 dB for the SRR methods (ℓ_2 and TV respectively), while with deep learning this improvement was higher, 21.04 dB with the subpixel and 22.02 dB with the U-net structure. If the SSI and IFC values ([0,1]) are considered as percentages, they improved compared to the CBCT by 1.59–2.01% and 15.54–15.81% respectively.

The PSNR value is the only metric where the U-net slightly outperforms the subpixel structure. As this metric uses the MSE, this fact relates to the previous result regarding the ℓ_{1s} loss function, where the U-net performs better than the subpixel structure. It shows that the subpixel CNN can grasp the inner structure of the image better, and the ℓ_1- and ℓ_{1s}-type losses training the networks are not directly the best measures for perceptually correct metrics.

3.3.4 Comparison of 3D segmented images

Table 3.2: Average Values of Canal Segmentation Metrics

Metric (against µCT)	CBCT	SRR:ℓ_2	SRR:TV	CNN:U-net	CNN:Subpixel
Area, mean difference (mm^2)	0.0510	0.0674	0.0634	0.0500	**0.0327**
Feret, mean difference (µm)	120.57	115.16	145.19	119.61	**114.26**
Difference of the canals (%)	12.39%	12.25%	12.40%	10.12%	**6.07**%
Dice coefficient	0.8891	0.8852	0.8913	0.8998	**0.9101**

The quantitative results of the segmentation can be seen in Table 3.2. The CBCT images and the results of the four enhancement methods were compared to the µCT images. In the table the averages of the absolute results on the four test teeth are shown.

The subpixel method clearly improved all the measures, which is most conspicouos with the difference of the volumes and mean of differences. The U-net gave better results too, but these were less considerable. The SRR techniques could slightly enhance some of the measures (see the Feret diameter for the SRR-ℓ_2, Dice coefficient for the SRR-TV method), but gave worse results than the CNN techniques.

As the quantitative results showed the subpixel method as the best technique, it was chosen for 3D-visualization. The segmented canal structures of the CBCT-µCT and subpixel-µCT volume pairs were compared. Fig. 3.8 shows three teeth from the test set with a color bar indicating the differences between the segmentation pairs. It can be seen that on the apical side of the root, where the diameter is smaller making the imaging and image segmentation more difficult, the deep learning technique estimated the structure more precisely. On the molar tooth a thinner lateral canal could be reconstructed.

Similarly to Section 3.3.2, where the performance metrics showed the CNN methods to be superior to SRR methods despite the lower contrast, the metrics here show that the segmentation was not affected by the lower contrast of the CNN methods.

3.4 Conclusion

In this chapter two different deep learning-based SISR methods were implemented for dental CBCT image enhancement. The techniques showed better results than state-of-the-art reconstruction-based SR approaches both in terms of quality metrics and subsequent image-segmentation-based analysis. It has been observed that the ℓ_{1s} loss function of the network is not directly the best measure for perceptually correct metrics like the SSI, IFC or PSNR. Even though the current method operates on slices of the volume, it offers a compatible solution in terms of speed and quality against the state of the art.

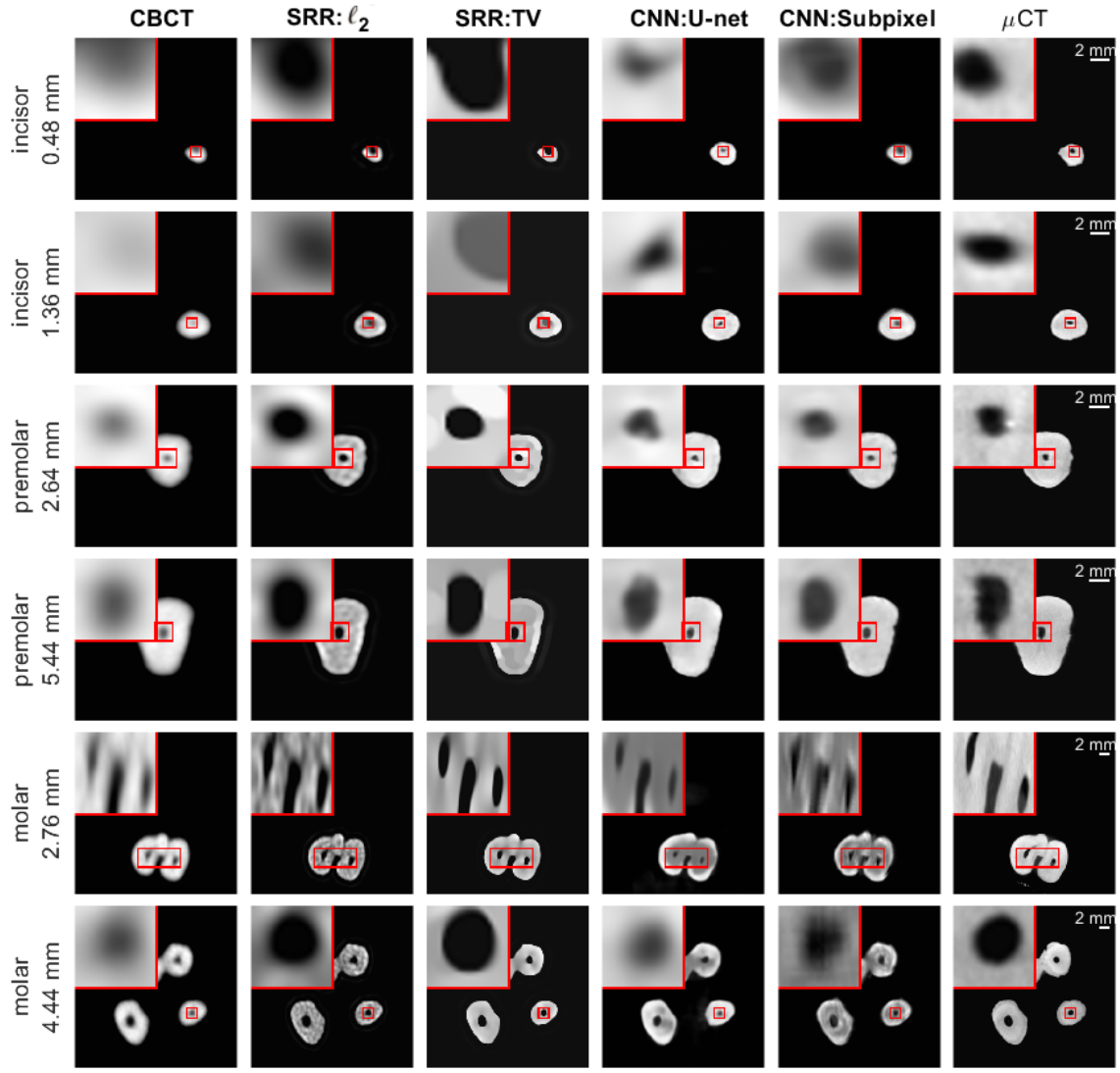

Figure 3.7: Result of SR methods on different slices from the test set. On the left of the first column the type of the tooth and the depth of the slice from the apex of the root is displayed. The columns stand for the four enhancement methods along with the original CBCT and µCT images. The enhancement was carried out after background-removal. It can be observed, that the SRR methods are tending to overestimate the size of the canal. In many cases the U-net shows a morphologically different shape. The result of the subpixel CNN is the most similar to the ground truth, as the metrics in Table 3.1 suggest. A 2 mm-scalebar is displayed on the µCT images. The display range is stretched to [0,1].

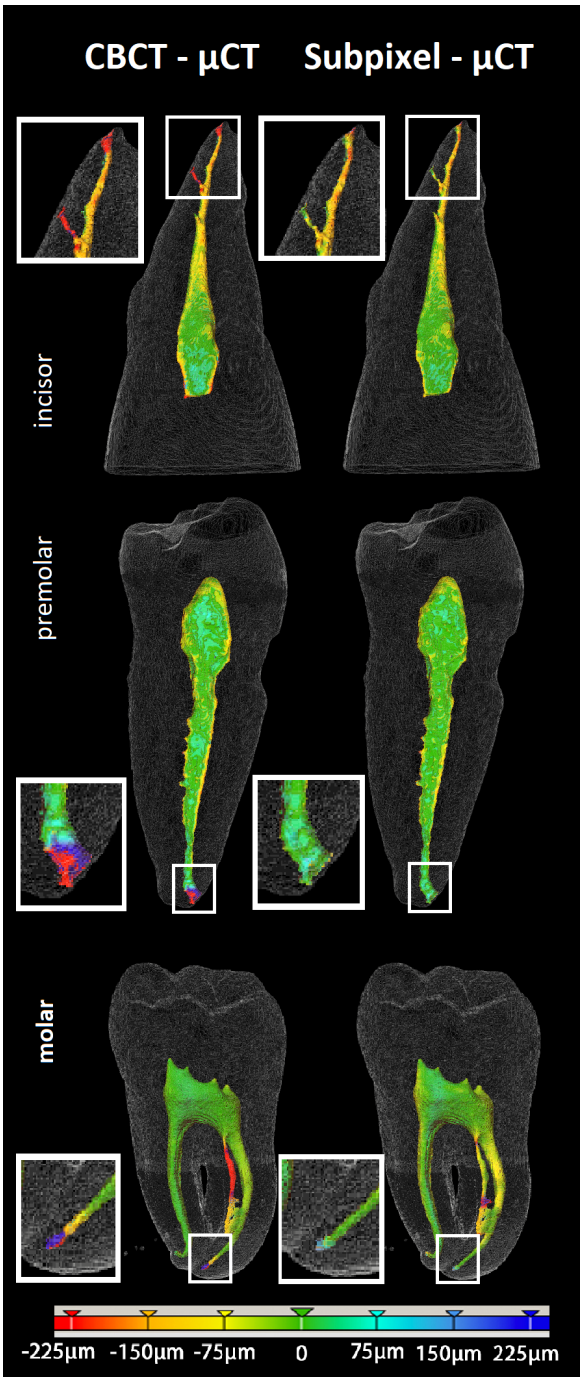

Figure 3.8: Volumetric segmentation of the root canal on the test set (an upper incisor, a lower premolar tooth and a lower molar). The colored area shows the difference between CBCT and µCT (on the left) and between the subpixel CNN and µCT segmentations. The highlighted areas show the apical end of the root, where the precision of the segmentation is more important during root canal treatment.

Chapter 4

Proposed tensor factorization - based image enhancement

In this chapter, 3D tensor-decomposition-based algorithms proposed during this PhD thesis for the SISR problem are introduced. First the CBCT image degradation model is defined for the µCT-CBCT images in form of a matrix-vector equation, which is used throughout the chapter. The first algorithm in Section 4.2 (written based on [J2]) uses the canonical polyadic decomposition for denoising, integrating it with the blurring and downsampling of the image in a single minimization term. This algorithm is extended with a semi-blind PSF estimation in Section 4.3, based on [C1]. In Section 4.4 the Tucker decomposition is applied in succession with a deblurring step to superresolve the CT images, based on the work in [C2]. All these methods show computationally lightweight solutions with qualitatively equivalent or better results compared to the state of the art introduced in Section 2.2.2.

4.1 The CBCT image degradation model

The image degradation model considered herein is the one classically used in SISR methods, which was introduced in Section 2.2.2, specifically in (2.4). The model relates the LR image (CBCT in the case of the current dental application) to an HR image (considered to be close to the µCT). The HR image $\boldsymbol{X} \in \mathbb{R}^{\mathrm{I}\times\mathrm{J}\times\mathrm{K}}$ is corrupted by decimation (with rate r), blurring, and some added noise

$\boldsymbol{N} \in \mathbb{R}^{\mathrm{I}/r \times \mathrm{J}/r \times \mathrm{K}/r}$, resulting in the LR image $\boldsymbol{Y} \in \mathbb{R}^{\mathrm{I}/r \times \mathrm{J}/r \times \mathrm{K}/r}$. The corresponding matrix-vector equation is

$$\mathrm{vec}(\boldsymbol{Y}) = DH\ \mathrm{vec}(\boldsymbol{X}) + \ \mathrm{vec}(\boldsymbol{N}). \tag{4.1}$$

We assume that $H \in \mathbb{R}^{\mathrm{IJK} \times \mathrm{IJK}}$ is the BCCB version of the 3D Gaussian kernel $\boldsymbol{H}$ to account for circular convolution. A 3D Gaussian kernel $\boldsymbol{H}$ is separable along the three dimensions to one-dimensional kernels as $\boldsymbol{H} = \mathbf{h}_1 \otimes \mathbf{h}_2 \otimes \mathbf{h}_3$ and is usually assumed for a blurring PSF [90]. For Gaussian convolutional arrays $\mathbf{h}_1, \mathbf{h}_2, \mathbf{h}_3$ with standard deviations $\overline{\sigma} = \{\sigma_1, \sigma_2, \sigma_3\}$ the corresponding BCCB matrices are $H_1 \in \mathbb{R}^{\mathrm{I} \times \mathrm{I}}, H_2 \in \mathbb{R}^{\mathrm{J} \times \mathrm{J}}, H_3 \in \mathbb{R}^{\mathrm{K} \times \mathrm{K}}$. The decimation matrix $D \in \mathbb{R}^{(\mathrm{I}/r \cdot \mathrm{J}/r \cdot \mathrm{K}/r) \times \mathrm{IJK}}$ downsamples the image by an integer number, by averaging blocks of r neighboring pixels in each direction. This matrix is also separable for the three dimensions as $D_1 \in \mathbb{R}^{\mathrm{I}/r \times \mathrm{I}}, D_2 \in \mathbb{R}^{\mathrm{J}/r \times \mathrm{J}}, D_3 \in \mathbb{R}^{\mathrm{K}/r \times \mathrm{K}}$. This formulation of D corresponds for instance to the physical process of a large CBCT detector element collecting rays over a larger area, than µCT does. This matrix also has better inversion properties compared to the regular decimation operator which discards pixels at a rate r.

4.2 CPD for 3D SISR

Available super-resolution techniques for 3D images are either computationally inefficient prior-knowledge-based iterative techniques or deep learning methods which require a large database of known low- and high-resolution image pairs. A recently introduced tensor-factorization-based data fusion approach offers a fast solution without the use of known image pairs or strict prior assumptions. In this work this factorization framework is investigated for single image resolution enhancement with an off-line estimate of the system PSF. For this purpose the 3D SISR minimization problem is formulated using the tensor CPD, and a computationally efficient analytical solution is given. The connection between the tensor rank and image complexity is also briefly investigated. The technique is applied to 3D CBCT dental image resolution enhancement. To demonstrate the efficiency of our method, it is compared to a recent state-of-the-art iterative technique using

low-rank and total variation regularizations [29]. In contrast to this comparative technique, the proposed reconstruction technique gives a 2-order-of-magnitude improvement in running time – 2 minutes compared to 2 hours for a dental volume of 282×266×392 voxels. Furthermore, it also offers slightly improved quantitative results (PSNR, segmentation quality). Another advantage of the presented technique is the low number of hyperparameters. As demonstrated in this work, the framework is not sensitive to small changes of its parameters, proposing an ease of use.

4.2.1 Problem formulation

The herein presented CPD-SISR algorithm follows the idea of [62], where CPD is used for the fusion of multi- and hyper-spectral images (detailed at the end of Section 2.3.2). It is different from the minimization problem of [62] in the following aspects. First, only one measured data volume is used here in contrast to the two measurements in the fusion problem. Second, here the degradation happens in all three dimensions between the HR and LR image, while in [62] the hyperspectral measurement is degraded in the first two dimensions, the multispectral volume in the third dimension.

As the real rank F of $\boldsymbol{X}$ is unknown, it will be estimated by R. Let $\hat{\overline{U}} = \left\{\hat{U}_1 \in \mathbb{R}^{\mathrm{I}\times R}, \hat{U}_2 \in \mathbb{R}^{\mathrm{J}\times R}, \hat{U}_3 \in \mathbb{R}^{\mathrm{K}\times R}\right\}$ be the estimated CPD of $\boldsymbol{X}$, following (2.27), (2.28), and (2.34). The image degradation problem can be rewritten following (2.29) using the separated kernel and decimation matrices as

$$\begin{aligned}\boldsymbol{Y} &= \boldsymbol{X} \times_1 D_1H_1 \times_2 D_2H_2 \times_3 D_3H_3 + \boldsymbol{N} \\ &\approx \hat{\boldsymbol{X}} \times_1 D_1H_1 \times_2 D_2H_2 \times_3 D_3H_3 + \boldsymbol{N} \\ &= [\![D_1H_1\hat{U}_1, D_2H_2\hat{U}_2, D_3H_3\hat{U}_3]\!] + \boldsymbol{N}.\end{aligned} \tag{4.2}$$

The SISR task can be defined as finding the set of matrices $\overline{U}$ that is the solution of the following minimization problem

$$\min_{\hat{\overline{U}}} \left\| \boldsymbol{Y} - [\![D_1H_1\hat{U}_1, D_2H_2\hat{U}_2, D_3H_3\hat{U}_3]\!] \right\|_F^2. \tag{4.3}$$

As problem (4.3) is NP-hard, an alternating optimization method is investigated for minimizing the cost function for $\hat{U}_1, \hat{U}_2, \hat{U}_3$, similarly to (2.31). Building a tensor

from its decomposition (2.27) consists of a summation of R outer products. Minimizing this sum for $\hat{U}_n$ would yield in $\frac{R(R-1)}{2} + R$ terms, leading to a complex cost-function. To understand this, remember that the square of the sum has to be taken, and $\left(\sum_{n=1}^{N} a_i\right)^2$ consists of $\frac{N(N-1)}{2}$ terms of $2a_i a_j$ (with $i \neq j$) and N terms of a_i^2.

Instead when minimizing for $\hat{U}_n$, the tensors are mode-n-matricized using (2.25) and (2.30) leading to

$$\begin{aligned}
&\min_{\hat{U}_1} \frac{1}{2} \left\| \boldsymbol{Y}^{(1)} - D_1 H_1 \hat{U}_1 (D_3 H_3 \hat{U}_3 \odot D_2 H_2 \hat{U}_2)^T \right\|_F^2 \\
&\min_{\hat{U}_2} \frac{1}{2} \left\| \boldsymbol{Y}^{(2)} - D_2 H_2 \hat{U}_2 (D_3 H_3 \hat{U}_3 \odot D_1 H_1 \hat{U}_1)^T \right\|_F^2 \\
&\min_{\hat{U}_3} \frac{1}{2} \left\| \boldsymbol{Y}^{(3)} - D_3 H_3 \hat{U}_3 (D_2 H_2 \hat{U}_2 \odot D_1 H_1 \hat{U}_1)^T \right\|_F^2 .
\end{aligned} \tag{4.4}$$

Note that the unfolding is performed in each direction sequentially, conserving the 3D local information. The three minimizations in (4.4) can be analytically solved using the least-square estimator, obtained by the Moore-Penrose pseudo-inverse as shown generally in (2.33). However, the inversion proved to be unstable on our dataset, and called for some regularization, namely a regularized Moore-Penrose pseudo-inverse (†) defined as

$$A^\dagger = (A^T A + \epsilon^2 I)^{-1} A^T, \tag{4.5}$$

where ϵ is a hyper-parameter used to provide a stable inverse (this procedure is classically referred to as Tikhonov regularization [91]). This can be mathematically derived by adding regularizers of the component norms to (4.3) as

$$\min_{\hat{U}} \left\| \boldsymbol{Y} - [\![D_1 H_1 \hat{U}_1, D_2 H_2 \hat{U}_2, D_3 H_3 \hat{U}_3]\!] \right\|_F^2 + \epsilon^2 \left(\left\| \hat{U}_1 \right\|^2 + \left\| \hat{U}_2 \right\|^2 + \left\| \hat{U}_3 \right\|^2 \right). \tag{4.6}$$

This can be unfolded similarly to (4.4) as

$$\begin{aligned}
&\min_{\hat{U}_1} \frac{1}{2} \left\| \boldsymbol{Y}^{(1)} - D_1 H_1 \hat{U}_1 (D_3 H_3 \hat{U}_3 \odot D_2 H_2 \hat{U}_2)^T \right\|_F^2 + \epsilon^2 \left\| \hat{U}_1 \right\|^2 \\
&\min_{\hat{U}_2} \frac{1}{2} \left\| \boldsymbol{Y}^{(2)} - D_2 H_2 \hat{U}_2 (D_3 H_3 \hat{U}_3 \odot D_1 H_1 \hat{U}_1)^T \right\|_F^2 + \epsilon^2 \left\| \hat{U}_2 \right\|^2 \\
&\min_{\hat{U}_3} \frac{1}{2} \left\| \boldsymbol{Y}^{(3)} - D_3 H_3 \hat{U}_3 (D_2 H_2 \hat{U}_2 \odot D_1 H_1 \hat{U}_1)^T \right\|_F^2 + \epsilon^2 \left\| \hat{U}_3 \right\|^2 .
\end{aligned} \tag{4.7}$$

For simplicity, the derivation is continued only with the first equation, with the substitution $Y = \boldsymbol{Y}^{(1)}$, $B = D_1 H_1$, $X = \hat{U}_1$, $C = (D_3 H_3 \hat{U}_3 \odot D_2 H_2 \hat{U}_2)^T$ as

$$\min_X \|Y - BXC\|_F^2 + \epsilon^2 \|X\|^2, \tag{4.8}$$

which after vectorization is

$$\min_{\mathrm{vec}(X)} \|\mathrm{vec}(Y) - \mathrm{vec}(BXC)\|_F^2 + \epsilon^2 \|\mathrm{vec}(X)\|^2. \tag{4.9}$$

The vectorization of BXC can be written using the generalized outer product (Kronecker product) as

$$\mathrm{vec}(BXC) = (C^T \otimes B)x \tag{4.10}$$

where $\mathrm{vec}(X) = x$. After substituting this into (4.9)

$$\min_x \left\| y - (C^T \otimes B)x \right\|^2 + \epsilon^2 \|x\|^2 \tag{4.11}$$

is obtained. The slope of the function is zero at the minimum, written as

$$\begin{aligned} \frac{\partial}{\partial x} \left\| y - (C^T \otimes B)x \right\|^2 + \epsilon^2 \|x\|^2 &= 0 \\ -2\left[(C^T \otimes B)^T y - (C^T \otimes B)^T (C^T \otimes B)x + \epsilon^2 x \right] &= 0 \\ \left[(C^T \otimes B)^T (C^T \otimes B) - \epsilon^2 \right]^{-1} (C^T \otimes B) y &= x. \end{aligned} \tag{4.12}$$

This is exactly the Tikhonov-regularized Moore-Penrose pseudo-inverse from (4.5) with $A = (C^T \otimes B)$, thus

$$(C^T \otimes B)^\dagger y = (C^{\dagger T} \otimes B^\dagger) y = x. \tag{4.13}$$

Using (4.10) it can be rewritten in the matrix form as

$$B^\dagger Y C^\dagger = X. \tag{4.14}$$

With back-substituting the values of B and C, and repeating these steps for the minimization of $\hat{U}_2$ and $\hat{U}_3$ from (4.7), the analytical solution implemented for (4.6) is

$$\begin{aligned} \hat{U}_1 &= (D_1 H_1)^\dagger \boldsymbol{Y}^{(1)} (D_3 H_3 \hat{U}_3 \odot D_2 H_2 \hat{U}_2)^{\dagger T} \\ \hat{U}_2 &= (D_2 H_2)^\dagger \boldsymbol{Y}^{(2)} (D_3 H_3 \hat{U}_3 \odot D_1 H_1 \hat{U}_1)^{\dagger T} \\ \hat{U}_3 &= (D_3 H_3)^\dagger \boldsymbol{Y}^{(3)} (D_2 H_2 \hat{U}_2 \odot D_1 H_1 \hat{U}_1)^{\dagger T}. \end{aligned} \tag{4.15}$$

Note that this standard inversion method provided good results for the application considered in this work. However, other solutions could be of interest and would deserve to be investigated in future work.

The computational complexity of the above solution is relatively low, as the pseudo-inverses (which is the bottleneck of the algorithm) require the inversion only of $I \times I$, $J \times J$, $K \times K$ and $R \times R$ matrices. For an $n \times n$ matrix the inversion has a complexity of $O(n^3)$ or somewhat lower, depending on the inversion algorithm. Throughout the algorithm $(D_n H_n)^\dagger$ does not change, their pseudo-inverse has to be calculated only once. However, $(D_n H_n \hat{U}_n \odot D_m H_m \hat{U}_m)^\dagger$ changes on each iteration, so three inversions with a complexity of $O(R^3)$ have to be calculated to update these values.

The proposed CPD-SISR method was implemented using some basic tensor functions from the TensorLab toolbox [92] in Matlab 2017b, namely the tensor structure, the Khatri-Rao product, the CPD initialization and the building of a tensor from its CPD. In the algorithm $\hat{\overline{U}}$ was initialized with elements from the standard normal distribution and $\hat{U}_1, \hat{U}_2, \hat{U}_3$ were updated iteratively several times as described in Algo. 1.

Algorithm 1 CPD-SISR algorithm [link to source code]

Input: $\boldsymbol{Y} \in \mathbb{R}^{\mathrm{I}/r \times \mathrm{J}/r \times \mathrm{K}/r}, R, [\sigma_1, \sigma_2, \sigma_3], r, \epsilon$

1: **Initialize** $\hat{\overline{U}} = \left\{ \hat{U}_1 \in \mathbb{R}^{\mathrm{I} \times F}, \hat{U}_2 \in \mathbb{R}^{\mathrm{J} \times F}, \hat{U}_3 \in \mathbb{R}^{\mathrm{K} \times F} \right\}$ with normally distributed values

2: $D_1, D_2, D_3 \leftarrow$ decimation operator with a factor r

3: $H_1, H_2, H_3 \leftarrow$ Gaussian kernels with standard deviations $[\sigma_1, \sigma_2, \sigma_3]$

4: **while** stopping criteria is not met **do**

5: $\quad \hat{U}_1 \leftarrow Y^{(1)}, \hat{U}_2, \hat{U}_3$

6: $\quad \hat{U}_2 \leftarrow Y^{(2)}, \hat{U}_1, \hat{U}_3$ (▷) update using (4.15)

7: $\quad \hat{U}_3 \leftarrow Y^{(3)}, \hat{U}_1, \hat{U}_2$

8: **end while**

9: $\hat{\boldsymbol{X}} \leftarrow \hat{\overline{U}}$ (▷) build using (2.27)

Output: $\hat{\boldsymbol{X}}$, the estimated high resolution image

---

4.2.2 Connection between the tensor rank and the image complexity

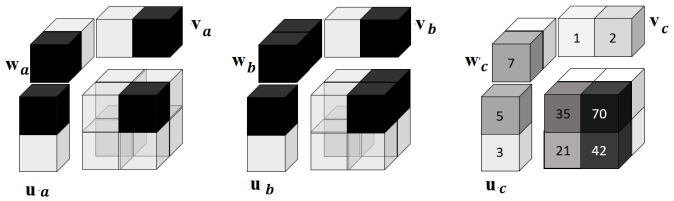

Figure 4.1: Tensor rank and image complexity. In example a) a single dark pixel (representing 1) in the white (representing 0) volume can be expressed by one outer product. In b) two neighboring pixels are dark, making one outer product sufficient for their description. In c) the pixel value is printed on the cell, equals 0 if not present. Two fibers are linearly dependent (2× [21,35] = [42,70]), so the volume can be decomposed using a tensor rank of $F = 1$.

In this work, an estimate factorization of R components is calculated, as the tensor rank F is NP-hard to obtain. In this section it is explained, how F is dependent on the complexity of the image, and how the number of components is related to the TV and low rank regularizers. The complexity of the image can be associated with piecewise constant volumes and dependent fibers of the image. To explain this claim, an illustrative set of examples is provided in Fig. 4.1. In these examples the notations and dimensionality of (2.23) are used with I = J = K = 2. Fig. 4.1 *a)* shows that the image with a single dark pixel (representing 1) in the white volume (representing 0) has a tensor rank of $F = 1$. More precisely

$$\begin{aligned} &\mathbf{u}_a \otimes \mathbf{v}_a \otimes \mathbf{w}_a = [1,0] \otimes [0,1] \otimes [1,0] = \boldsymbol{X_a} \\ &\boldsymbol{X}_{\boldsymbol{a}}^{(3)} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}. \end{aligned} \tag{4.16}$$

In Fig. 4.1 *b)*, two neighboring pixels are dark. This does not change the complexity of the image, since one outer product can still describe this volume. Indeed, we have

$$\begin{aligned} &\mathbf{u}_b \otimes \mathbf{v}_b \otimes \mathbf{w}_b = [1,0] \otimes [0,1] \otimes [1,1] = \boldsymbol{X_b} \\ &\boldsymbol{X}_{\boldsymbol{b}}^{(3)} = \begin{bmatrix} 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}. \end{aligned} \tag{4.17}$$

In Fig. 4.1 *c)*, two fibers are linearly dependent with 2×[21,35]=[42,70], which also makes one outer product sufficient for decomposing the tensor since

$$\begin{aligned} \mathbf{u}_c \otimes \mathbf{v}_c \otimes \mathbf{w}_c &= [5,3] \otimes [1,2] \otimes [7,0] = \boldsymbol{X_c} \\ \boldsymbol{X}_{\boldsymbol{c}}^{(3)} &= \begin{bmatrix} 35 & 70 & 0 & 0 \\ 21 & 42 & 0 & 0 \end{bmatrix}. \end{aligned} \tag{4.18}$$

This set of illustrative examples shows that for images with piecewise constant regions (like the neighboring cells in Fig. 4.1 *b)*) or with low matrix rank (as the linearly dependent fibers in Fig. 4.1 *c)*) a smaller tensor rank can be expected. More generally, the tensor decomposition (2.27) tends to promote solutions with small tensor ranks. This property is useful in the case of denoising, when independent outlier pixels have to be eliminated. A degraded image may contain larger constant areas, with higher dependency between neighboring rows and columns. It means that describing these images will also be more efficient with a tensor of small rank. Thus these simple examples allow us to understand why CBCT images can be represented by a reduced number of rank-1 tensors, allowing identifiability of the decomposition. This train of thought allows an insight into how Algo. 1 operates. A small R allows the denoising of the volume with some low rank and TV properties on the output, while the D and H matrices realize the deblurring and upsampling operations. As these are incorporated into a joint minimization, the algorithm is expected to be even more advantageous.

4.2.3 Results and discussion

Comparison to an existing 3D SISR method

The state-of-the-art LRTV introduced in Section 2.2.2 was used as a benchmark to compare the performance of the proposed method. The parameters used for testing can be seen in Table 4.1. They were tuned manually to get the highest possible improvement of the PSNR. The tests were run on a standard personal computer (PC) with an Intel(R) Core(TM) i7 2×2.5GHz processor and 16 GB of RAM.

Table 4.1: Parameters

LRTV	CPD-SISR
$n_{\text{ADMM}} = 5$	$n_{\text{TF}} = 10$
$\sigma = [5.8, 5.3, 0.9]$	$\sigma = [5.8, 5.3, 0.9]$
$\lambda_{TV} = 0.02$	$R = 500$
$\lambda_R = 0.05$	$\epsilon = 1$
$\rho = 0.05$	
$n_{grad} = 100$	
$dt = 0.05$	

The two methods were tested for three samples from the dataset $\overline{\boldsymbol{TE}}_{\boldsymbol{L}}$, including an incisor, a premolar and a molar. The sizes of the sample volumes, the PSNR calculated against the µCT images from $\overline{\boldsymbol{TE}}_{\boldsymbol{H}}$ and the execution times are provided in Table 4.2. Compared to the CBCT images the PSNR improves similarly for the LRTV (average of 1.2 dB) and the CPD-SISR (average of 1.5 dB) methods with the chosen parameters. However, this enhancement is achieved at a much lower computational cost: 10 iterations of CPD-SISR run 100 times faster than 5 iterations of LRTV. This faster execution time is important since it permits a wider range of applications, including those requiring a rapid diagnosis during a medical examination. In Fig. 4.2, the quality of the enhanced volumes is visualized, showing that the canal is better defined and contrasted compared to the CBCT image, suggesting better segmentation properties.

For further analysis the root canal was segmented from each volume, using the segmentation method described in Section 3.1.3. Qualitative and quantitative results are provided in Fig. 4.3 and in Table 4.3.

In particular, Table 4.3 shows differences between the estimates and the values obtained using the µCT image for three parameters (Feret diameter, area of the canal and Dice coefficient). The estimated Feret diameter improves similarly with both SR techniques compared to the CBCT images with an averaged improvement of 63 µm for LRTV and 81 µm for CPD-SISR. The second line of the table shows how the area of the canal on the axial slices is changing from one method to another.

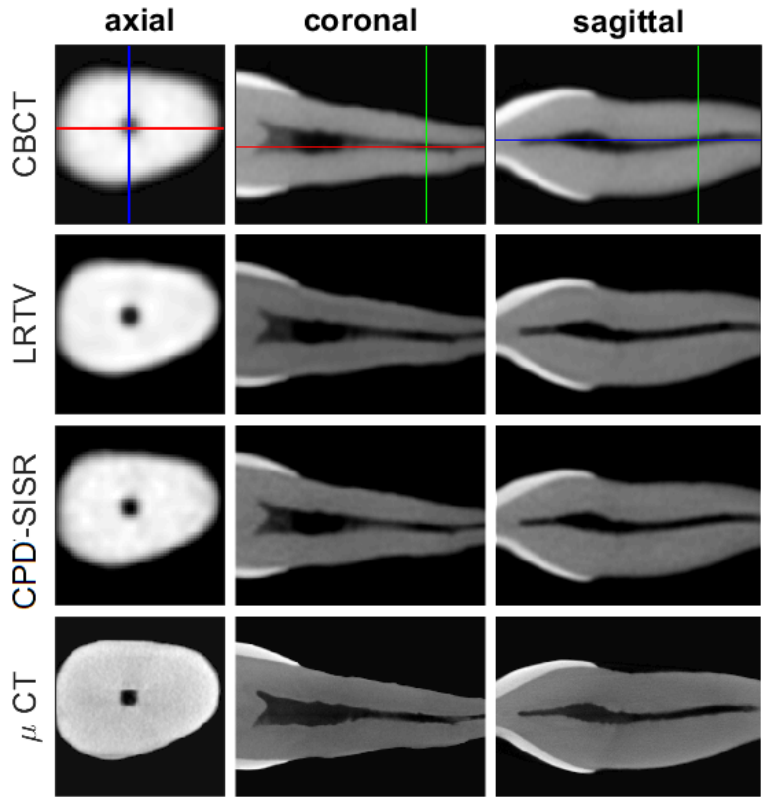

Figure 4.2: Results on Sample #1. In the rows the CBCT, LRTV output, CPD-SISR output and μCT images can be seen, whereas the columns correspond to one axial, a coronal and a sagittal slice. The CBCT image is shown at the higher scale of the HR images, for better comparison. The location of the slices within the volume is illustrated on the CBCT images in colored lines.

Note that the LRTV method shows a higher difference compared to the original CBCT (by 0.0256 mm^2), suggesting that the TV regularization overestimates the canal. This observation is also confirmed in Fig. 4.3, as the LRTV volumes have a more blueish color corresponding to positive differences. The CPD-SISR provides the best overall performance with an improvement of 0.0152 mm^2 on average. The last metric in the table is the Dice coefficient, also showing some improvement in the overlap of the canals, by 1% using the LRTV and 2% with the CPD-SISR. Fig. 4.3 displays zoomed-in sections of the apical part of the canal, as this part is the most important during the treatment. Considering these results, the CPD-SISR method shows slightly better segmentation properties than the LRTV technique, while offering a great reduction in running time.

Table 4.2: Test results

	Sample #1	Sample #2	Sample #3
tooth type	upper incisor	lower premolar	lower molar
µCT image size	282×266×392	280×268×492	324×306×402
CBCT PSNR	23.17 dB	22.67 dB	24.14 dB
LRTV PSNR	**24.32** dB	**24.65** dB	24.61 dB
CPD-SISR PSNR	**24.32** dB	24.48 dB	**25.71** dB
LRTV time	6988 s	9059 s	10301 s
CPD-SISR time	**71 s**	**92 s**	**104 s**

Table 4.3: Canal Segmentation Metrics

	method	Sample #1	Sample #2	Sample #3	mean
Mean of Diff. - Feret (µm)	CBCT	96	89	341	176
	LRTV	74	71	196	113
	CPD-SISR	**50**	**57**	**177**	**95**
Mean of Diff. - Area (mm^2)	CBCT	0.0463	0.0461	0.2492	0.1139
	LRTV	0.0914	0.0920	0.2350	0.1395
	CPD-SISR	**0.0447**	**0.0271**	**0.2243**	**0.0987**
Dice coefficient	CBCT	0.88	0.88	0.90	0.88
	LRTV	0.87	0.88	0.90	0.89
	CPD-SISR	**0.90**	**0.91**	**0.91**	**0.90**

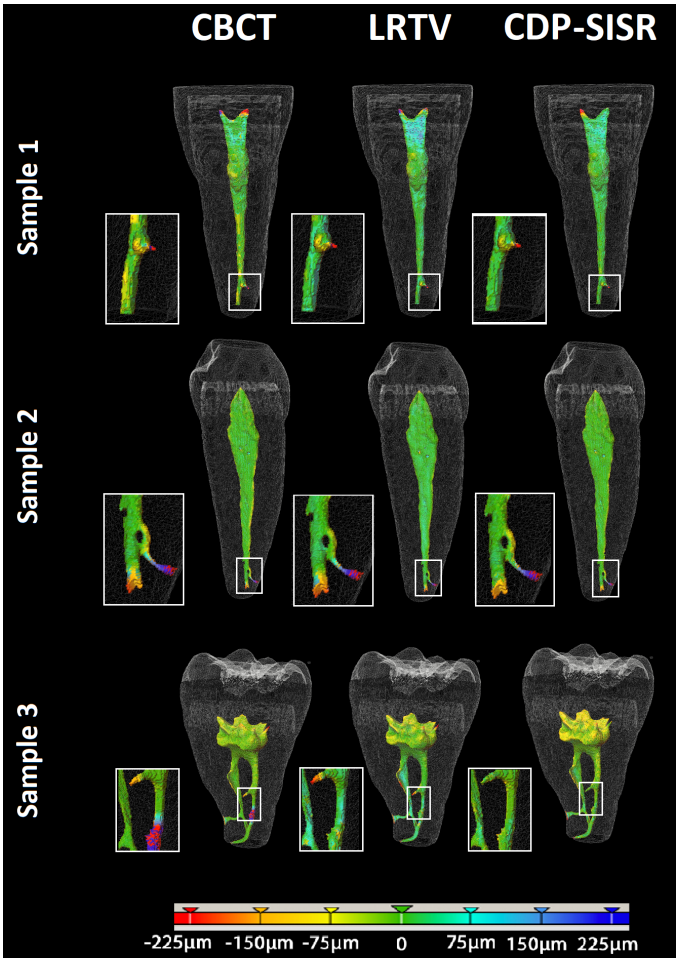

Figure 4.3: Segmentation results for CBCT, LRTV and CPD-SISR for the 3 samples. The color-bar visualizes the distance between the estimated surface of the canal and the one obtained with μCT segmentation.

Adjusting the parameters of the CPD-SISR method

The impact of the tensor rank and the iteration number was investigated using Sample #1. Fig. 4.5 *a)* shows that the runtime increases linearly with the number of iterations, as expected. Fig. 4.5 *a)* also shows that the PSNR converges rapidly to its maximum value (close to 24.5), which is an interesting property of the proposed method. Fig. 4.4 *a)* shows how the solutions qualitatively evolve with the iteration number. For improved visibility the difference from the $n_{\mathrm{TF}} = 10$ case is shown in the figure. In the case of $n_{\mathrm{TF}} = 10$ a second test was run, and the difference was calculated compared to this result, as the random initialization of $\hat{\overline{U}}$ results in slightly different outputs. It can be seen that as the iteration number increases, the

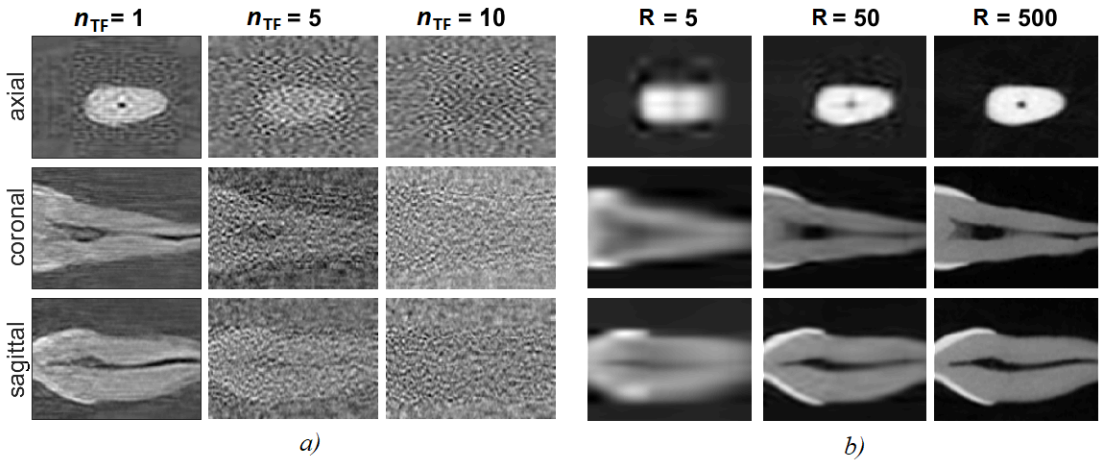

Figure 4.4: Effect of the iteration number it and the estimated tensor rank R in the reconstructed images. Sample #1 is visualized through 3 slices from the axial, coronal and sagittal directions. In panel a) the difference compared to a result obtained after 10 iterations is shown. In case of n_{TF}=10, a second test run was used for calculating the difference (note that the algorithm has random initialization, therefore different runs result in slightly different outputs). With more iterations the difference becomes less structured, more random. In panel b) the change with R can be seen: low numbers cause large blocks in the images, and the higher the tensor rank, the more detailed the output is.

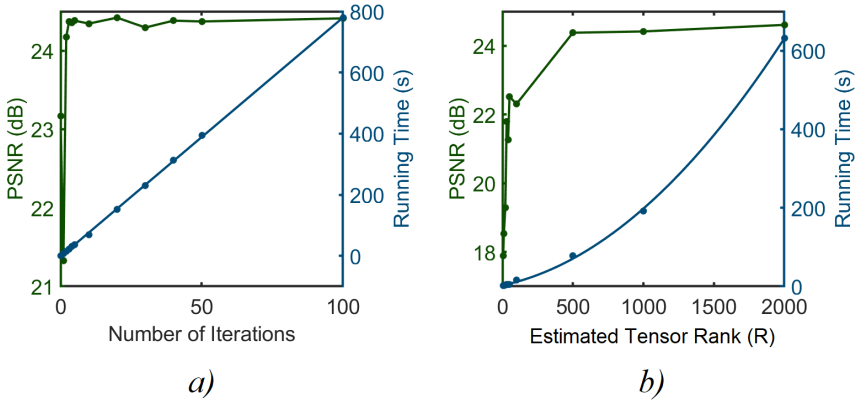

Figure 4.5: Effect of the iteration number and the estimated tensor rank on the PSNR and runtime. The rest of the parameters are as in Table 4.1. In a) the PSNR saturates after a small number of iterations, while the runtime increases linearly. In b) the runtime has am exponential growth versus the estimated tensor rank and the PSNR saturates around $R = 500$.

difference becomes less structured: in the third column the shape of the tooth is almost invisible and it is lost in random noise.

The tensor rank F was estimated by R. According to the upper limit on F for a unique CPD, in the case of Sample #1, $R \leq 2^{14}$ should be efficient. However, numbers higher than 2000 caused memory problems, and were therefore not tested. Fig. 4.5 *b)* shows that the computational time increases exponentially with R since the algorithm requires the inversion of larger matrices in $\hat{\overline{U}}$. It can also be seen that the PSNR stabilizes for $R \geq 500$. Some sample images can be seen in Fig. 4.4 *b)* showing that low values of R lead to large constant blocks in the image, which is characteristic of a low-rank or TV regularization. For larger numbers finer details become visible.

Note that our results indicate that neither of the above parameters have to be estimated precisely. After a small number of iterations the result converges. R can be considered as a prior information on the complexity of the image. Using higher values a more natural result can be obtained, but above a threshold the method will not give more precise outputs.

4.3 CPD for 3D SISR with semi-blind PSF-estimation

A volumetric non-blind single image super-resolution technique using CPD has been introduced in 4.2. In this section, a joint alternating recovery of the high-resolution image – using the previously introduced CPD-SISR technique,– and of the unknown PSF parameters – assuming a Gaussian function with standard deviations from a pre-defined interval – is proposed. The method is evaluated on dental computed tomography images. The algorithm was again compared to the LRTV 3D SISR, now combined with the same alternating PSF-optimization. The two algorithms have shown similar improvement in PSNR, but our method converged roughly 40 times faster, under 6 minutes both in simulation and on experimental dental computed tomography data.

4.3.1 Problem formulation

Let $\hat{\overline{U}} = \left\{\hat{U}_1 \in \mathbb{R}^{\mathrm{I}\times F}, \hat{U}_2 \in \mathbb{R}^{\mathrm{J}\times F}, \hat{U}_3 \in \mathbb{R}^{\mathrm{K}\times F}\right\}$ be the CPD of HR tensor $\boldsymbol{X}$. Using (2.29) and the separated kernels, (4.1) can be written as

$$\begin{aligned}\boldsymbol{Y} &= \boldsymbol{X} \times_1 D_1H_1 \times_2 D_2H_2 \times_3 D_3H_3 + \boldsymbol{N} \\ &= [\![D_1H_1\hat{U}_1, D_2H_2\hat{U}_2, D_3H_3\hat{U}_3]\!] + \boldsymbol{N}.\end{aligned} \tag{4.19}$$

The semi-blind SISR task is now to find the set of matrices $\hat{\overline{U}}$ and parameters $\overline{\sigma} = \{\sigma_1, \sigma_2, \sigma_3\}$ which minimizes

$$\min_{\hat{\overline{U}},\overline{\sigma}} \left\| \boldsymbol{Y} - [\![D_1H_1(\sigma_1)\hat{U}_1, D_2H_2(\sigma_2)\hat{U}_2, D_3H_3(\sigma_3)\hat{U}_3]\!] \right\|_F^2. \tag{4.20}$$

As problem (4.20) is non-convex, an alternating optimization method is proposed, minimizing (4.20) for $\hat{U}_1, \hat{U}_2, \hat{U}_3$ and $\sigma_1, \sigma_2, \sigma_3$, respectively.

Image estimation

The image estimation step is calculated as described in Section 4.2. Compared to (4.4) the only difference is that the blurring BCCB matrices are recalculated in each iteration from $\overline{\sigma}$.

$$\begin{aligned}&\min_{\hat{U}_1} \frac{1}{2} \left\| \boldsymbol{Y}^{(1)} - D_1H_1(\sigma_1)\hat{U}_1(D_3H_3(\sigma_3)\hat{U}_3 \odot D_2H_2(\sigma_2)\hat{U}_2)^T \right\|_F^2 \\ &\min_{\hat{U}_2} \frac{1}{2} \left\| \boldsymbol{Y}^{(2)} - D_2H_2(\sigma_2)\hat{U}_2(D_3H_3(\sigma_3)\hat{U}_3 \odot D_1H_1(\sigma_1)\hat{U}_1)^T \right\|_F^2 \\ &\min_{\hat{U}_3} \frac{1}{2} \left\| \boldsymbol{Y}^{(3)} - D_3H_3(\sigma_3)\hat{U}_3(D_2H_2(\sigma_2)\hat{U}_2 \odot D_1H_1(\sigma_1)\hat{U}_1)^T \right\|_F^2.\end{aligned} \tag{4.21}$$

The least-square estimator of (4.21) obtained with the Moore-Penrose pseudo-inverse and Tikhonov regularization, with the recalculated blurring matrices is

$$\begin{aligned}\hat{U}_1 &= (D_1H_1(\sigma_1))^\dagger \boldsymbol{Y}^{(1)}(D_3H_3(\sigma_3)\hat{U}_3 \odot D_2H_2(\sigma_2)\hat{U}_2)^{\dagger T} \\ \hat{U}_2 &= (D_2H_2(\sigma_2))^\dagger \boldsymbol{Y}^{(2)}(D_3H_3(\sigma_3)\hat{U}_3 \odot D_1H_1(\sigma_1)\hat{U}_1)^{\dagger T} \\ \hat{U}_3 &= (D_3H_3(\sigma_3))^\dagger \boldsymbol{Y}^{(3)}(D_2H_2(\sigma_2)\hat{U}_2 \odot D_1H_1(\sigma_1)\hat{U}_1)^{\dagger T}.\end{aligned} \tag{4.22}$$

PSF estimation

The PSF estimation was implemented based on [93]. As H is a BCCB matrix, H vec($\mathbf{X}$) of (4.1) can be rewritten using the Fourier transform $\mathcal{F}$ in its 3D version

as

$$H \text{ vec}(\mathbf{X}) = \mathcal{F}^{-1}(\mathcal{F}\tilde{\boldsymbol{H}} \circ \mathcal{F}\mathbf{X}) \tag{4.23}$$

where $\tilde{\boldsymbol{H}}$ is the zero-padded, circularly shifted version of kernel $\boldsymbol{H}$ and the operation ($\circ$) is the element-wise matrix multiplication (Hadamard product). This formulation radically reduces the size of the matrices to be multiplied. The kernel $\tilde{\boldsymbol{H}}$ can be expressed as

$$\tilde{\boldsymbol{H}}(\overline{\sigma}) = \frac{1}{\sigma_1\sigma_2\sigma_3\sqrt{2\pi}^3} e^{\frac{1}{2}\left(\frac{x^2}{\sigma_1}+\frac{y^2}{\sigma_2}+\frac{z^2}{\sigma_3}\right)}, \tag{4.24}$$

where x, y, z are the fixed 3D evaluation coordinates of the zero-padded, shifted kernel. It can be assumed, based on prior knowledge depending on the application, that all elements of $\overline{\sigma}$ are within a given interval. This is expressed by an indicator function

$$\mathbf{1}_{[a,b]}(\overline{\sigma}) = \mathbf{1}_{[a_1,b_1]}(\sigma_1) + \mathbf{1}_{[a_2,b_2]}(\sigma_2) + \mathbf{1}_{[a_3,b_3]}(\sigma_3), \tag{4.25}$$

where

$$\mathbf{1}_{[a_n,b_n]}(\sigma_n) = \begin{cases} 0 & \text{if} \quad \sigma_n \in [a_n, b_n] \\ \beta & \text{if} \quad \sigma_n \notin [a_n, b_n] \end{cases} \tag{4.26}$$

with $n \in \{1, 2, 3\}$ and penalty β. Combining this with (4.23) and (4.24) the kernel optimization problem can now be written as

$$\min_{\overline{\sigma}} \left\| \boldsymbol{Y} - \mathcal{F}^{-1}(\mathcal{F}\tilde{\boldsymbol{H}}(\overline{\sigma}) \circ \mathcal{F}\mathbf{X}) \right\|_F^2 + \mathbf{1}_{[a,b]}(\overline{\sigma}). \tag{4.27}$$

Introducing a function $G(\overline{\sigma})$ for the first, data fidelity term of (4.27), its compact form becomes

$$\min_{\overline{\sigma}} G(\overline{\sigma}) + \mathbf{1}_{[a,b]}(\overline{\sigma}). \tag{4.28}$$

Knowing that the solution of the proximal operator for the above problem is a projection ($\prod$) onto the corresponding $[a_1, b_1], [a_2, b_2], [a_3, b_3]$ intervals, the solution of (4.27) will be the fixed point of

$$\overline{\sigma} = \prod \left(\overline{\sigma} - \gamma \frac{dG(\overline{\sigma})}{d\overline{\sigma}} \right), \tag{4.29}$$

where γ, the step size is a small enough coefficient (its value might be changed at each iteration). In (4.29) the data fidelity term $G(\overline{\sigma})$ is estimated with a gradient descent step and the indicator function is taking effect through the projection operator.

The CPD-SISR-blind algorithm

The image and kernel estimation steps were joined, denoted by CPD-SISR-blind and shown in Algorithm 2. The two problems are minimized in an alternating manner. The algorithm was implemented in Matlab 2017b, and for the basic tensor operations and the tensor structure the TensorLab toolbox was employed [92].

Algorithm 2 CPD-SISR-blind algorithm [link to source code]

Input: $\boldsymbol{Y}, R, a_1, b_1, a_2, b_2, a_3, b_3\overline{\sigma}_0, r, N, M, \epsilon$

Initialize $\hat{\overline{U}}$ with normally distributed values
$D_1, D_2, D_3 \leftarrow$ decimation operator with a factor r
for i = 0:N **do**
    $H_1, H_2, H_3 \leftarrow$ Gaussian kernels of $\overline{\sigma}_i$
    for j = 0:M **do**
        $\hat{U}_1 \leftarrow Y^{(1)}, \hat{U}_2, \hat{U}_3$
        $\hat{U}_2 \leftarrow Y^{(2)}, \hat{U}_1, \hat{U}_3$ (▷) update using (4.22)
        $\hat{U}_3 \leftarrow Y^{(3)}, \hat{U}_1, \hat{U}_2$
    end for
    $\hat{\mathbf{X}} \leftarrow \hat{\overline{U}}$ (▷) build using (2.27)
    Initialize diff= inf
    while diff $> \epsilon$ **do**
        $\overline{\sigma}^{i+1} \leftarrow \overline{\sigma}^{i}, \hat{\mathbf{X}}$ (▷) update using (4.29)
        diff = $\max(\overline{\sigma}^{i+1} - \overline{\sigma}^{i})$
    end while
end for

Output: $\boldsymbol{X}$, the estimated high resolution image

4.3.2 Results and discussion

Within the simulation the LR images were obtained by blurring and downsampling the µCT images (considered as the ground truth in CT imaging) without added noise, with parameters listed in Table 4.4. In the experimental setting, the LR input was the CBCT volume and its µCT pair was the ground truth HR volume.

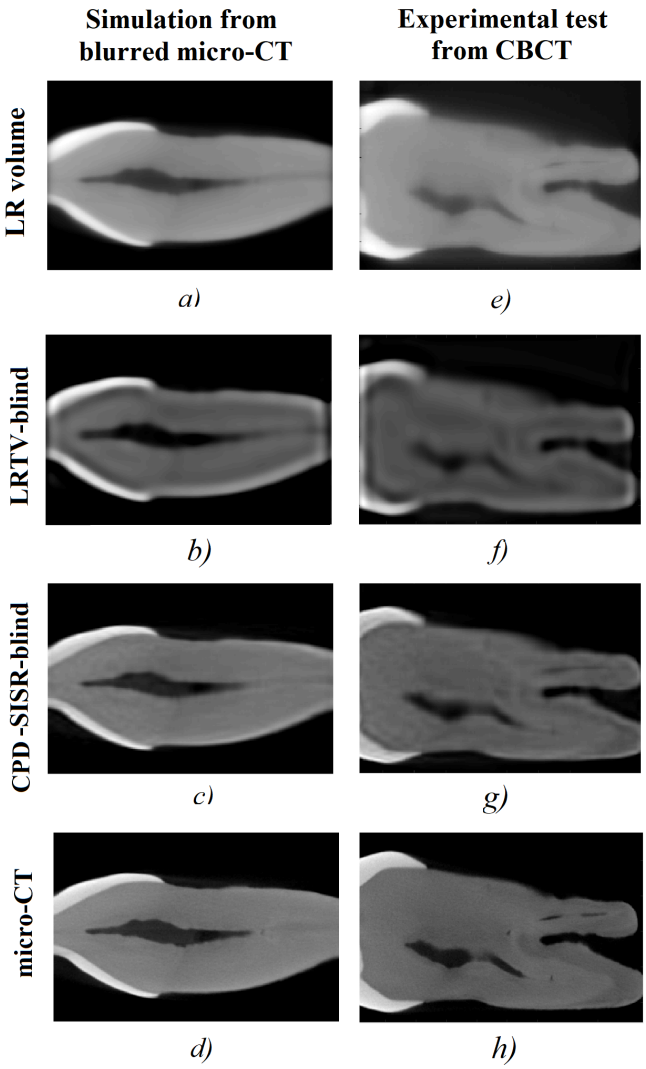

Figure 4.6: Image enhancement. On the left slices corresponding to the simulation, on the right slices from the real-data volume can be seen. A coronal slice was chosen for demonstration. The LR images are shown on the scale of the HR images using linear interpolation.

The proposed method was again compared with the LRTV method described in Section 2.2.2. The original code was combined with the PSF estimation step of (4.29). The method is further denoted as LRTV-blind. The CPD-SISR-blind algorithm ran 3 overall iterations ($N = 3$), 5 image-update iterations in each loop ($M = 5$), and the kernel update repeated until convergence. The initial step size γ was set to 1e-5 and decreased linearly in the outer loop. The parameter setting for the simulation and experimental data is listed in Table 4.4. The LRTV-blind algorithm ran 3 overall iterations, in each loop running 3 LRTV iterations for image update, followed by Gaussian parameter optimization until convergence. The hy-

Table 4.4: CPD-SISR-blind parameters and quantitative results

	Simulation	Experiment
HR pixel number	287×266×392	274×278×474
chosen R	400	
downsampling rate r	2	
ground truth $\overline{\sigma}$	[6.0 6.0 6.0]	–
initialized $\overline{\sigma}$	[8.0 8.0 7.0]	[8.0 8.0 7.0]
$\overline{\sigma}$ with LRTV-blind	[4.7 4.6 **6.3**]	[7.6 6.5 7.4]
$\overline{\sigma}$ with CPD-SISR-blind	[**5.0** **4.9** 4.8]	[8.5 7.8 6.5]
LR–HR PSNR	22.32 dB	19.42 dB
LRTV-blind PSNR	24.39 dB	25.63 dB
CPD-SISR-blind PSNR	**26.53** dB	**30.07** dB
LRTV-blind time	9087 s	11823 s
CPD-SISR-blind time	**298 s**	**354 s**

perparameters controlling the regularization were set to 0.07 for the low-rank and 0.02 for total variation. Note that for both methods, all the hyperparameters were manually tuned to provide the best results possible in terms of visual inspection. In both methods the initialization of $\overline{\sigma}$ is started from the values listed in Table 4.4, and the $[a_1, b_1], [a_2, b_2], [a_3, b_3]$ intervals of the projections are set corresponding to the initialized $\overline{\sigma}$ as $[\overline{\sigma} - 4, \overline{\sigma} + 2]$. For the experimental data no ground truth value of the Gaussian parameters was available.

The results are illustrated through one slice extracted from the volumes in Fig. 4.6. It can be seen that with LRTV-blind the image is contrasted, smoothed and the root canal (dark region inside the tooth) is more dilated as a result of the TV regularization, both in the simulation and on the experimental data. Note that no such parameters need to be tuned in CPD-SISR-blind, resulting in more natural images.

In Table 4.4 the quantitative results are listed. The enhancement was measured quantitatively through the PSNR between the ground truth µCT and the LR, CPD-

SISR-blind, LRTV-blind images respectively. In simulation the PSNR improved by 18.9% in CPD-SISR and 9.3% in LRTV, while on the experimental data by 54.8% and 31.9% respectively. It can be seen that the Gaussian parameters of the simulation converged to a lower value than used for blurring, because of the non-convex nature of the problem. However, the most important improvement of CPD-SISR-blind remains its runtime, being roughly 40 times faster compared to the LRTV-blind on a standard PC with an Intel® Core™ i7 2×2.5GHz processor and 16 GB of RAM.

4.4 TD for 3D SISR

Compared to the previously introduced CPD-SISR, here the added value brought by Tucker decomposition was investigated. While CPD allows a joint implementation of the denoising and deconvolution steps of the SISR model, with Tucker decomposition the denoising is realized first, followed by deconvolution. This way the ill-posedness of the deconvolution caused by noise is partially mitigated. The results achieved using the two different tensor decomposition techniques were compared, and the robustness against noise was investigated. For validation the dental images were used. The superiority of the proposed method is shown in terms of PSNR, SSI, the canal segmentation accuracy, and runtime.

4.4.1 Problem formulation

The idea of the method proposed herein is to denoise the image before deblurring, in order to stabilize the deconvolution operation, as earlier suggested in the literature [94]. In this work [C2], the blurring kernel is assumed to be known, and is estimated for the current application as explained in Section 3.1.2 as

As explained in Section 2.3.3, the SV_n of each mode (SV_1, SV_2, SV_3) can be calculated from $\mathbf{\Sigma}$. Similarly to the 2D case, by picking the relevant components having a singular value higher than a threshold R_n, a denoised version of $\boldsymbol{Y}$, $\hat{\boldsymbol{Y}}$ may

be achieved [95].

$$\begin{aligned} \hat{\boldsymbol{Y}} &= \tilde{\boldsymbol{\Sigma}} \times_1 \tilde{V_1} \times_2 \tilde{V_2} \times_3 \tilde{V_3} \\ \text{where} \\ \tilde{V_n} &= V_n(:, i) \quad | \quad SV_n(i) \geq R_n \\ \tilde{\boldsymbol{\Sigma}} &= \boldsymbol{Y} \times_1 \tilde{V_1}^T \times_2 \tilde{V_2}^T \times_3 \tilde{V_3}^T \end{aligned} \tag{4.30}$$

Unlike in the 2D case, this truncated approximation might not be optimal in the least squares sense, but gives a reasonable estimate [63].

After obtaining the denoised image, the deconvolution and upsampling is realized using a Tikhonov-regularized deconvolution separated for the three modes.

$$\hat{\boldsymbol{X}} = \hat{\boldsymbol{Y}} \times_1 (D_1 H_1)^\dagger \times_2 (D_2 H_2)^\dagger \times_3 (D_3 H_3)^\dagger. \tag{4.31}$$

The algoritm is summarized in Algo. 3. The bottleneck of the computational complexity in this solution is the calculation of the three SVDs for the unfolded tensor for obtaining the basis matrices. This represents the calculation of SVDs for $IJ \times K$, $IK \times J$ and $JK \times I$ matrices, with $O(IJK^2)$, $O(IJ^2K)$ and $O(I^2JK)$ complexity, which is higher compared to the complexity of CPD-SISR. However, it has to be calculated only once, making this more efficient for smaller tensors, but more dependent on the image size. It is also worth mentioning that fast SVD algorithms do exist, which could accelerate both TD-SISR and CPD-SISR in future research [96, 97].

4.4.2 Results and discussion

Simulation results

In simulation the µCT scan of a lower premolar was chosen (280×268×492 pixels). It was blurred with a Gaussian kernel (standard deviation σ_1=σ_2=σ_3=8), downsampled (at r=2) and Gaussian noise was added (white noise at different SNR levels) for its LR counterpart.

For CPD-SISR $R = 500$ was chosen, and in both methods $\epsilon = 1$ was set for the Tikhonov-regularization (4.5) following [J2]. As it can be seen in Fig. 4.7, the singu-

Algorithm 3 TD-SISR algorithm [link to source code]

Input: $\boldsymbol{Y} \in \mathbb{R}^{\mathrm{I}/r \times \mathrm{J}/r \times \mathrm{K}/r}, R_1, R_2, R_3, [\sigma_1, \sigma_2, \sigma_3], r, \epsilon$

Denoising block

1: $V_n, \boldsymbol{\Sigma} \leftarrow \boldsymbol{Y} = [\![\Sigma; V_1, V_2, V_3]\!]$ (▷) using (2.35) and (2.36)

2: $SV_n \leftarrow \boldsymbol{\Sigma}$ (▷) calculate using (2.38)

3: $\hat{\boldsymbol{Y}} \leftarrow \boldsymbol{\Sigma}, V_n, SV_n, R_n$ (▷) denoise using (4.30)

Deconvolution block

4: $D_1, D_2, D_3 \leftarrow$ decimation operator with a factor r

5: $H_1, H_2, H_3 \leftarrow$ Gaussian kernels with $[\sigma_1, \sigma_2, \sigma_3]$

6: $\hat{\boldsymbol{X}} \leftarrow \hat{\boldsymbol{Y}}, D_n, H_n$ (▷) deconvolve using (4.31)

Output: $\hat{\boldsymbol{X}}$, the estimated high resolution image

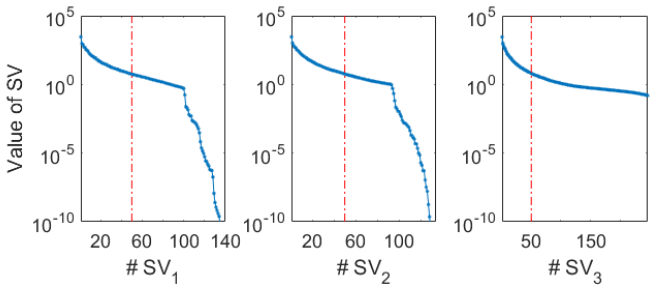

Figure 4.7: Simulation - singular values without added noise in all three modes, on a logarithmic scale. The vertical lines represent the chosen truncation thresholds.

lar values decay rapidly (mind the logarithmic scale). For TD-SISR R_1=R_2=R_3=40 (SV_n values under 1) were chosen as they were generally sufficient for all noise levels.

Table 4.5 shows the quantitative results of the SISR methods. The PSNR is improved for each case and both methods, compared to the simulated LR image. TD-SISR gave better results, except for the extremely noisy, 20 dB case. The SSI gave similar results, with no improvement in the 20 dB case. After these results the segmentation was carried out at 25 dB. The improvement is confirmed by the Dice coefficients, showing the superiority of the TD-SISR method.

Real data results

For real data experiments an upper molar (324×248×442 pixels) was used. The LR images were the CBCT volumes, with estimated standard deviations of σ_1=8.2,

Table 4.5: Metrics in simulation

	Simulated LR	CPD-SISR	TD-SISR
runtime	-	17.96 s	**1.86 s**
	PSNR (dB)		
no noise	28.56	31.48	**34.99**
30 dB	28.45	31.17	**34.39**
25 dB	28.36	31.08	**31.40**
20 dB	27.98	**30.01**	29.33
	SSI [0, 1]		
no noise	0.9623	0.9680	**0.9823**
30 dB	0.9612	0.9650	**0.9763**
25 dB	0.9572	0.9595	**0.9653**
20 dB	0.**9463**	0.9453	0.9417
	Segmentation at 25 dB		
Dice	0.8976	0.9242	**0.9425**

σ_2=7.5, σ_3=1.3. The volumes naturally contain noise because of the measurements and of the reconstruction algorithm. However, the level of noise is low for extracted teeth due to the absence of surrounding structures. To mitigate this and to further explore the robustness of the SISR algorithms to noise, Gaussian noise corresponding to different SNRs was artificially added to the experimental data.

For the real data in CPD-SISR the same settings were used as in simulation, and for TD-SISR $R_1, R_2, R_3 = 50$ was set after plotting the SVs (Fig. 4.8).

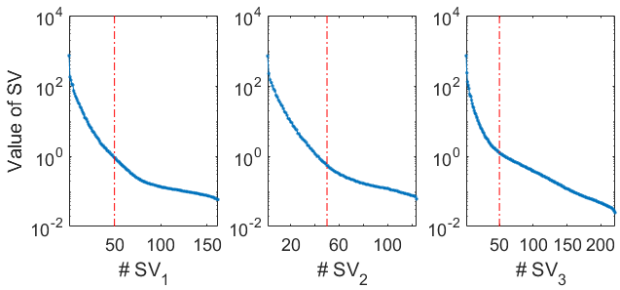

Figure 4.8: Real data - singular values without added noise in all three modes, on logarithmic scale. The vertical lines represent the chosen truncation thresholds.

The metrics have shown milder improvements compared to the simulation, but both PSNR and SSI improved in all cases with both methods, and TD-SISR gave superior results. The volume of the segmented 25 dB images also improved regarding the Dice coefficient.

Table 4.6: Metrics in real data

	CBCT	CPD-SISR	TD-SISR
runtime	-	17.71 s	**1.46 s**
	PSNR (dB)		
no noise	19.55	21.25	**21.61**
30 dB	19.30	20.84	**21.57**
25 dB	19.10	20.13	**21.09**
20 dB	18.91	20.21	**20.29**
	SSI [0, 1]		
no noise	0.8647	0.8907	**0.8935**
30 dB	0.8610	0.8870	**0.8929**
25 dB	0.8478	0.8784	**0.8908**
20 dB	0.8173	0.8555	**0.8814**
	Segmentation at 25 dB		
Dice	0.8939	0.9189	**0.9304**

Discussion

In contrast to the earlier CPD-SISR method no iterations are applicable in TD-SISR. In TD-SISR 3 thresholds R_1, R_2, R_3 have to be defined for the three modes, while in CPD-SISR only one parameter, R influences the denoising step. However, the singular values of TD-SISR correspond to the importance of the components, while R in CPD-SISR bears no such meaning. This makes the setting of TD-SISR parameters easier, and its efficiency is validated by the qualitative and quantitative results. The runtime of TD-SISR is lower because of the lack of iterations, but calculating the SVD for even larger volumes might be a bottleneck [98].

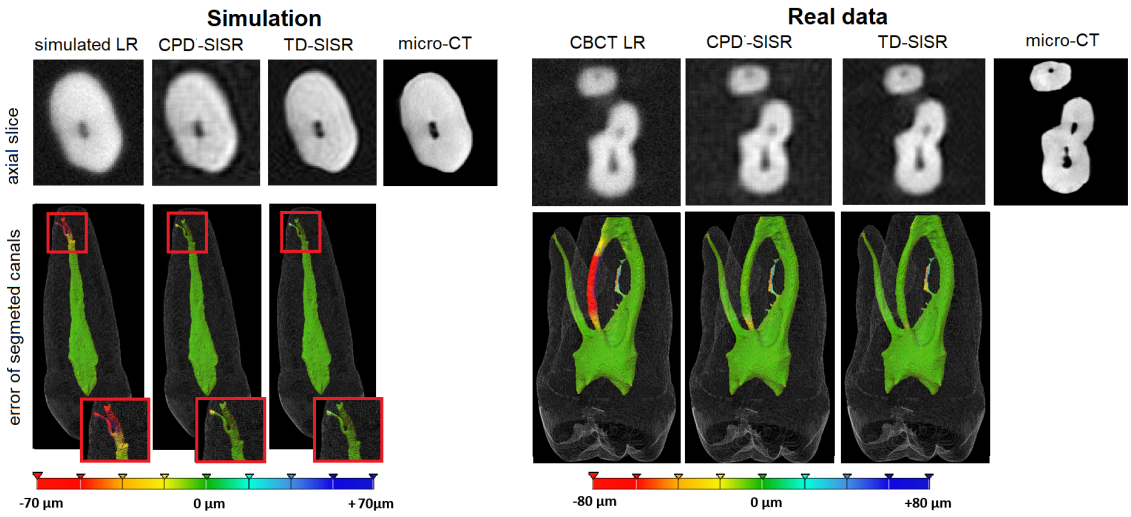

Figure 4.9: Results of SISR methods under 25 dB noise, both in simulation and in real data. The first row shows a single axial slice taken from the volumes. The second row shows the distance between the segmented HR and LR, enhanced LR volumes.

4.5 Conclusion

In this chapter three contributions of this thesis were introduced, using tensor factorization for solving the 3D SISR problem.

First a CPD-based method was proposed. I have embedded the SISR degradation model into CPD, and provided the corresponding minimization-problem. I gave a solution for the recovery of the HR image in the form of a compact ALS algorithm, using Tikhonov-regularization. This method is showing interesting computational advantages compared to currently available regularization-based methods, with slightly improved image quality compared to the investigated LRTV technique. The runtime of this method was about 100 times faster than with LRTV, allowing a wider field of applications. The method also uses significantly less parameters (tensor rank and iteration number) that can be easily adjusted by visual inspection of the reconstruction results. Dental CBCT volumes used as experimental data showed higher PSNR and canal-segmentation values, with similar or improved results compared to the LRTV method. Considering these results, the method was found to be promising for 3D SISR.

A possible direction for improving the CPD-SISR algorithm is to embed a PSF-

estimation into the framework. In the second proposed algorithm, I have introduced a fast method for SISR with joint Gaussian kernel estimation. First I have formulated the extended cost function of the problem, then implemented an iterative solution, alternatingly minimizing for the HR image, and for the Gaussian parameters. The algorithm managed to improve the image quality similarly to an existing ADMM reconstruction method using LR and TV regularization. The main advantage of the algorithm is its speed, processing 3D volumes of 287×266×392 and 274×278×474 pixels in less than 5 minutes with standard Matlab implementation on a PC (with Intel(R) Core(TM) i7 2×2.5GHz processor and 16 GB of RAM).

As CPD is not the only tensor factorization technique, a fair question is whether CPD is the best choice. I have found that the Tucker decomposition may be used for denoising the data. The framework of this decomposition did not allow embedding the degradation model in it, but proved to be efficient in denoising the image volume. This gave the idea of denoising the tensor before applying deconvolution. Thus, in the third algorithm, I have proposed a new SISR technique, using Tucker decomposition for the denoising, and a Tikhonov-regularized deconvolution for the deblurring step. Even though two additional parameters have to be set, it gave faster and quantitatively better results in noisy images compared to CPD-SISR. Images of 280×268×492 and 324×248×442 pixels were super-resolved under 2 s with standard Matlab implementation, on the same PC.

Chapter 5

Discussion

5.1 Answers to the posed questions

In this thesis the 3D SISR problem was investigated, to enhance the quality of medical images, in particular that of dental CT scans. This is necessary for precise mapping of the dental cavity prior to root canal treatments, as the success rate of such procedures is only 60-85% in general practice.

It was shown that deep learning structures, the U-net and subpixel networks, can be applied with success for enhancing the resolution of CBCT slices. Even though the size of the data set and restriction of the GPU memory did not allow training on 3D images, the quality of the slices (Table 3.1), and the precision of the 3D canal segmentation (Table 3.2) are both improvable by the presented algorithms. The improvement is most prominent in the apical region (Fig. 3.8), where treatments usually fail. To our knowledge, this was the first application of deep learning super-resolution in the field of dental images, and was followed by new researches like [99–101].

Two different frameworks were designed using tensor decompositions to give solution to the 3D SISR problem. This way the enhancement profits from the 3D information of the structure, while state-of-the-art methods discard a dimension when they unfold the volume. The algorithm using CPD decomposition updates the solution in an alternating manner for the three dimensions, realizing simultaneous denoising and deblurring of the image. The method improved the overall image

and segmentation qualities (Tables 4.2 and 4.3). The strength of this algorithm is its speed, which depends on the dimensional separability of the blurring and downsampling operators. In case of a tensor $\boldsymbol{X} \in \mathbb{R}^{\mathrm{I}\times\mathrm{J}\times\mathrm{K}}$ it allows computations with small matrices (e.g. $H_1 \in \mathbb{R}^{\mathrm{I}\times\mathrm{I}}$) compared to the generally used large system matrices of unfolded volumes ($H \in \mathbb{R}^{\mathrm{IJK}\times\mathrm{IJK}}$), allowing two-orders-of-magnitude faster execution times. Furthermore, the algorithm proved to be robust to its parameters (Fig. 4.5).

In the second framework it is shown that the TD can also be successfully applied to the posed problem. In this case the denoising of noisy images is realized by the decomposition step, followed by an efficient regularized deconvolution of the volume. This solution is a generalized version of the 2D truncated SVD, offering a semi-optimal (in the least squares sense), but fast solution - its runtime was an order of magnitude less, than with the CPD decomposition, and offered improved quality metrics both in simulation (Table 4.5) and on real data (Table 4.6).

Finally, a framework was constructed, where the PSF of the system and the HR images are estimated jointly. In the previous tensor-based algorithms the PSF was supposed to be known, and is estimated from the whole dataset. One not always has the means to obtain such an estimate, therefore blind solutions are desired. The standard deviations of the Gaussian kernel are supposed to be in a known interval, hence the algorithm is semi-blind. The image is enhanced with the CPD-SISR algorithm, offering an alternating optimization of the two terms. The algorithm managed to improve the LR images giving good estimates of the HR images, both in simulation and on real data (Table 4.4). Note that this blind deconvolution technique is similar to deep learning techniques, as neither of them requires a predefined PSF. However, while CPD-SISR-blind estimates the system function from a single LR image, the neural networks learn the PSF implicitly from a set of LR-HR image pairs.

Tensor implementation proved to be a strong tool in 3D SISR problems. The main advantage of these algorithms is their speed, as it allows their diagnostic use in real life settings. While state-of-the art algorithms super-resolve dental CBCT scans (approximately 300×300×400 pixels) in hours, CPD-SISR does so in minutes,

and TD in seconds, in Matlab on the same PC. Possible tracks for their further investigation are introduced in the following section.

5.2 Perspectives

The work discussed in this thesis opened many possible paths for future research. We have already started investigating some of these possibilities, others are loose ends waiting to be continued.

5.2.1 Deep learning

As deep learning is gaining growing interest among researchers, further development possibilities are published on a weekly (if not daily) basis.

In the proposed technique the most interesting direction would be an *in vivo* application. A transient step toward this goal is the implantation of the extracted teeth into a phantom jaw [102], adding realistic noise, tissue background to the CBCT scans. Such images were already collected, and for the increased database it was necessary to rewrite the algorithm to read and augment the data on-the-fly. Recently the possibility was investigated whether the network could be trained directly for a segmented HR output.

Even though neither the size of the dataset, nor the current hardware allows the application of real 3D CNNs for this problem, the information retrieved from neighboring slices could be crucial for obtaining a continuous dental cavity. As a compromise, multiple slices could be used as the input of the network, similarly to the color channels of [54].

We have concluded, that the chosen loss functions do not correspond to the perceptual metrics applied. In this regard the efficiency of different loss-functions and adversarial networks [103] could be investigated.

5.2.2 Tensor factorization

Tensor factorization is also a popular topic, more and more papers dealing with higher dimensional data in the form of tensors are being published.

In future work, exploring the domain of medical images that are of low tensor rank (thus are identifiable using CPD-SISR or CPD-SISR-blind), and investigating the influence of tensor rank on image recovery could be interesting. The prior information in regularization-based techniques is often empirical and guides the solution. Future work can investigate if such classical priors could be included in these frameworks, and whether they would improve the result. Also, the connection between the CPD-SISR and TD-SISR parameters along with their robustness, and the general inverse problem including thresholding constraint can be investigated.

The application of the techniques on different modalities is also of interest. CPD-SISR was already applied with success on MRI images in [104], where its speed is acknowledged.

Chapter 6

Summary

In this chapter the new scientific results are summarized in the form of thesis points in English, French and Hungarian.

6.1 New scientific results

The resolution of dental cone beam computed tomography (CBCT) images is limited by detector geometry, sensitivity, patient movement, the reconstruction technique and the need to minimize radiation dose. The corresponding image degradation model assumes that the CBCT image is a blurred (with a point spread function, PSF), downsampled, noisy version of a high resolution image. The quality of the image is crucial for precise diagnosis and treatment planning. The methods proposed in this thesis aim to give a solution for the single image super-resolution (SISR) problem. The algorithms were evaluated on dental CBCT and corresponding high-resolution (and high radiation-dose) µCT image pairs of extracted teeth.

Thesis I: *I have designed a deep learning framework for the SISR problem, applied to CBCT slices. I have tested the U-net and subpixel neural networks, which both improved the PSNR by 21-22 dB, and the Dice coefficient of the canal segmentation by 1-2.2%, more significantly in the medically critical apical region.*
Corresponding publication: [J1]

Convolutional neural networks have shown promising results for resolution en-

hancement [54]. To our knowledge this was the first time that a deep learning algorithm was used for biomedical SISR.

The U-net network [61] allows feature extraction on five different scales and combines their information on the output. My implementation used batch normalization for generalization, leaky rectified linear unit activation for avoiding inactive neurons, and a modified Hubert-loss for more accurate training. The subpixel network [58] extracts features directly from the low resolution image through six layers, and realizes the upsampling with a depth-to-space tiling operation in the last layer. It offers a computationally lightweight, still efficient solution for SISR.

CBCT – µCT image pairs of 5680 axial slices taken from 13 teeth were used for training, and 1824 slices of 4 teeth for testing the networks. Two existing reconstruction-based super-resolution methods using ℓ_2-norm and total variation regularization were used for comparison. The results were evaluated with different metrics, as the peak signal-to-noise ratio, structural similarity index, and subsequent 3D image-segmentation-based analysis.

The results show the superiority of the proposed CNN-based approaches over the state of the art in the case of dental CT images, allowing better detection of medically salient features such as the size, shape, or curvature of the root canal. It has been observed that the‘ chosen loss function of the network is not directly the best measure for perceptually correct metrics, as they only moderately affirm the visually observed enhancement.

Thesis II a: *I have designed an algorithm for the 3D SISR problem, using the canonical polyadic decomposition of tensors. This implementation conserves the 3D structure of the volume, integrating the factorization-based denoising, deblurring with a known PSF, and upsampling of the image in a lightweight algorithm with a low number of parameters. It outperforms the state-of-the-art 3D reconstruction-based algorithms with two orders of magnitude faster run-time and provides similar PSNR (improvement of 1.2-1.5 dB) and segmentation metrics (Dice coefficient increased on average to 0.89 and 0.90).*
Corresponding publication: [J2]

The canonical polyadic decomposition (CPD) of 3D tensors has recently been used for the fusion of multi- and hyperspectral images [62]. CPD finds the smallest set of pure tensors (outer product of three arrays), which sums up to the tensor in question. In case a smaller set is used, a denoised tensor may be expressed.

The proposed CPD-SISR algorithm optimizes for the set of pure tensors, which composes the denoised, upscaled, deblurred (with an estimated PSF) version of the CBCT volume, and does so in a fused implementation only alternating among the dimensions. The main advantage compared to the state of the art lies in the tensor-implementation, avoiding the formulation of large, $X \in \mathbb{R}^{IJK \times IJK}$ matrices from $\boldsymbol{X} \in \mathbb{R}^{I \times J \times K}$ tensors, still preserving the 3D information.

The results were compared to a state-of-the-art, reconstruction-based algorithm with total variation and low rank regularization, LRTV. Because of the large matrices this method is computationally extremely heavy, enhancing a sample volume of $282 \times 266 \times 392$ pixels in two hours, raising difficulties in the tuning of its six parameters. The proposed algorithm executed for the same volume in a little over a minute, using only three robust parameters. The PSNR increased similarly for the two methods, while the segmentation was significantly better in case of CPD-SISR. These results were promising enough for further research, as described in the following thesis points.

Thesis II b: *I have implemented a joint alternating recovery of the unknown PSF parameters and of the high-resolution 3D image using CPD-SISR. The algorithm was compared to a state-of-the-art 3D reconstruction-based algorithm, combined with the proposed alternating PSF-optimization. The two algorithms have shown similar improvement in PSNR, but CPD-SISR-blind converged roughly 40 times faster, under 6 minutes both in simulation and on experimental dental computed tomography data.*

Corresponding publication: [C1]

For the direct estimation of the PSF a dataset of known low- and high-resolution image pairs, or dedicated measurements on a phantom are necessary, repeated for any machinery of which the output images are to be enhanced. Otherwise the PSF

has to be estimated along with the deblurred image in a joint manner.

In this work a semi-blind estimation was realized, assuming that the standard deviations of the Gaussian PSF are within a known interval. The problem optimizing for these parameters can be solved with gradient descent. This minimization for the PSF and the CPD-SISR optimizing for the high-resolution image are repeated alternating until convergence.

For comparison the LRTV algorithm was used as in Thesis II a, combined with the proposed PSF-estimation in a similar alternating manner, denoted by LRTV-blind. In simulation the PSNR improved by 18.9% in CPD-SISR and 9.3% in LRTV-blind, while on the experimental data by 54.8% and 31.9% respectively. However, the most important improvement of CPD-SISR-blind remains its runtime, being roughly 40 times faster compared to the LRTV-blind.

Thesis II c: *I have proposed a solution for the 3D SISR problem using the Tucker decomposition (TD-SISR). The denoising step is realized first by TD in order to mitigate the ill-posedness of the subsequent deconvolution. Compared to CPD-SISR the algorithm runs ten times faster. Depending on the amount of noise, higher PSNR (0.3 - 3.5 dB), SSI (0.58 - 2.43%) and segmentation values (Dice coefficient, 2% improvement) were measured. The parameters in TD-SISR are familiar from 2D SVD-based algorithms, so their tuning is easier compared to CPD-SISR.*
Corresponding publication: [C2]

TD is the higher order generalization of the 2D singular value decomposition. The basis vectors may be organized according to their importance in the factorization. While the CPD defines a single rank that has to be estimated, TD uses the n-rank, three different values in case of 3D tensors. By thresholding the singular values with these estimated ranks, a denoised tensor can be composed. Here the deblurring can not be incorporated into the factorization algorithm, therefore they are implemented subsequently.

Even though two additional parameters have to be set, it gave faster and quantitatively better results in noisy images compared to the previous method, CPD-SISR. Images of 280×268×492 and 324×248×442 pixels were super-resolved under

2 s with standard Matlab implementation. The PSNR is improved under 20, 25, 30 dB added noise in both methods. TD-SISR outperformed CPD-SISR both in PSNR and SSI values, except for the extremely noisy, 20 dB case. After these results the segmentation was carried out at 25 dB. The improvement is confirmed by the Dice coefficients, showing the superiority of the TD-SISR method.

6.2 Contributions de la thèse

La résolution spatiale des images acquises par tomographie volumique à faisceau conique (CBCT) est limitée par la géométrie des capteurs, leur sensibilité, les mouvements du patient, les techniques de reconstruction d'images et la limitation de la dose de rayonnement. Le modèle de dégradation d'image considéré dans cette thèse consiste en un opérateur de flou avec la fonction d'étalement du système d'imagerie (PSF), un opérateur de décimation, et du bruit, qui relient les volumes CBCT à une image 3D super-résolue à estimer. Les méthodes proposées dans cette thèse (SISR - single image super-resolution) ont comme objectif d'inverser ce modèle direct, c'est à dire d'estimer un volume haute résolution à partir d'une image CBCT. Les algorithmes ont été évalués dans le cadre d'une application dentaire, avec comme vérité terrain les images haute résolution acquises par micro CT (µCT), qui utilise des doses de rayonnement très importantes, incompatibles avec les applications cliniques.

Contribution I *Nous avons proposé une approche de SISR par deep learning, appliquée individuellement à des coupes CBCT. Deux types de réseaux ont été évalués : U-net et subpixel. Les deux ont amélioré les volumes CBCT, avec un gain en PSNR de 21 à 22 dB et en coefficient de Dice pour la segmentation canalaire de 1 à 2.2%. Le gain a été plus particulièrement important dans la partie apicale des dents, ce qui représente un résultat important étant donnée son importance pour les applications cliniques.*
Publication associée : [J1]

Contribution II-a *Nous avons proposé des algorithmes de SISR basés sur la décomposition canonique polyadique des tenseurs. Le principal avantage de cette méthode, lié à l'utilisation de la théorie des tenseur, est d'utiliser la structure 3D des volumes CBCT. L'algorithme proposé regroupe plusieurs étapes: débruitage basé sur la factorisation des tenseurs, déconvolution et super-résolution, avec un faible nombre d'hyperparamètres. Le temps d'exécution est très faible par rapport aux algorithmes existants (deux ordres de magnitude plus petit), pour des performances légèrement supérieures (gain de 1.2 à 1.5 dB en PSNR).*
Publication associée : [J2]

Contribution II-b *La troisième contribution de la thèse est en lien avec la contribution 2: l'algorithme de SISR basé sur la décomposition canonique polyadique des tenseurs est combiné avec une méthode d'estimation de la PSF, inconnues dans les applications pratiques. L'algorithme résultant effectue les deux tâche de manière alternée, et s'avère précis et rapide sur des données de simulation et expérimentales.*
Publication associée : [C1]

Contribution II-c *La dernière contribution de la thèse a été d'évaluer l'intérêt d'un autre type de décomposition tensorielle, la décomposition de Tucker, dans le cadre d'un algorithme de SISR. Avant la déconvolution, le volume CBCT est débruité en tronquant sa décomposition de Tucker. Comparé à l'algorithme de la contribution 2, cette approche permet de diminuer encore plus le temps de calcul, d'un facteur 10, pour des performances similaires pour des SNR importants et légèrement supérieures pour de faibles SNR. Le lien entre cette méthode et les algorithmes 2D basés sur une SVD facilite le réglage des hyperparamètres comparé à la décomposition canonique polyadique.*
Publication associée : [C2]

6.3 Új tudományos eredmények

A fogászati cone beam komptertomográfiával (CBCT) készült felvételek felbontását a detektor geometriája, szenzitivitása, a páciens mozgása, a képrekonstrukciós technika, és a sugárdózis szükségszerű minimalizása befolyásolja. A folyamatot leíró degradációs modell a CBCT felvételre úgy tekint, mint a magas felbontású tárgy alulmintavételezett, zajos, elmosódott (a rendszer szóróválasz függvényével, PSF) párjára. A kezelések tervezéséhez és a pontos diagnózis érdekében elengedhetetlen a jó minőségű kép. A jelen disszertációban közölt módszerek célja megoldást nyújtani az egy felvételen alapuló szuper-rezolúciós problémára (SISR). Az algoritmusokat eltávolított fogak CBCT felvételén validáltam, amikhez a magas felbontású (és sugárdózisú) képpárt mikro-CT (µCT) felvételek szolgáltatták.

I. Tézispont: *Az SISR probléma megoldására egy mélytanulási módszeren alapuló keretrendszert terveztem, melyet fogászati CBCT szeletekre alkalmaztam. Az U-net és subpixel típusú neurális hálókat teszteltem, melyek mindketten javították a CBCT képek csúcs jel-zaj viszonyát (PSNR) 21-22 dB-lel, és a fogcsatorna szegmentációjának Dice-koefficiensét 1-2.2%-kal – hangsúlyosabban az orvosilag kritikus gyökércsúcsnál.*
Kapcsolódó publikáció: [J1]

II. Tézispont – a.: *A 3D SISR probléma megoldására algoritmust terveztem, mely a tenzorok kanonikus poliadikus felbontását (CPD) használja. Ez az implementáció megőrzi a 3D objektum strukturális információját, egy költég-hatékony és kevés paramétert használó algoritmusban ötvözve a tenzorfelbontáson alapuló zajszűrést, a PSF dekonvolúcióját és a kép felbontásának növelését. A korszerű rekonstrukció-alapú algoritmusokhoz hasonló eredményeket két nagyságrenddel gyorsabban ér el – a PSNR 1.2 -1.5 dB-lel, a szegmentáció Dice-koefficiense 0.89-0.90-re nőtt).*
Kapcsolódó publikáció: [J2]

II. Tézispot – b: *Egy algoritmusban egyesítettem a 3D kép CPD-SISR-alapú javítását és az ismeretlen PSF paraméterbecslését. A módszert egy korszerű rekonstrukció-alaú, 3D-SISR módszerhez hasonlítottam, melyet az általam javasolt PSF-becsléssel egészítettem ki. A módszerek hasonló javulást mutattak a PSNR szempontjából mint szimulált, mind valós CBCT képeken, de a javasolt módszer körülbelül 40-szer gyorsabban konvergált, kevesebb, mint 6 perc alatt.*

Kapcsolódó publikáció: [C1]

II. Tézispont – c: *Javaslatot tettem a 3D-SISR probléma megoldására a Tucker-dekompozíció alkalmazásával (TD-SISR). Először a kép zajszűrését végzem el a Tucker-dekompozíció segítségével, így javítható az ezt követő dekonvolúció rosszul kondícionáltsága. A CPD-SISR módszerhez képest 10-szer gyorsabban fut az algoritmus. A zaj függvényében növekedik a PSNR (0.3 -3.5 dB) és az SSI (0.58 - 2.43%), és a szegmentáció pontosságát mérő Dice-koefficiens (2%). Az algoritmusban használt paraméterek a 2D SVD-alapú zajszűrő algoritmusokéhoz hasonlóak, így könnyebben állíthatóak a CPD-SISR-hez képest.*

Kapcsolódó publikáció: [C2]

Journal publications of the thesis

[J1] J. Hatvani, A. Horváth, J. Michetti, A. Basarab, D. Kouamé, and M. Gyöngy, "Deep learning-based super-resolution applied to dental computed tomography," *IEEE Transactions on Radiation and Plasma Medical Sciences*, vol. 3, no. 2, pp. 120–128, 2018. (Cited on page(s): xii, 14, 20, 32, 82, 86, 88)

[J2] J. Hatvani, A. Basarab, J.-Y. Tourneret, M. Gyöngy, and D. Kouamé, "A tensor factorization method for 3-d super resolution with application to dental ct," *IEEE transactions on Medical Imaging*, vol. 38, no. 6, pp. 1524–1531, 2018. (Cited on page(s): xii, 14, 27, 52, 72, 83, 87, 88)

Conference publications of the thesis

[C1] J. Hatvani, A. Basarab, J. Michetti, M. Gyöngy, and D. Kouamé, "Tensor-factorization-based 3D single image super-resolution with semi-blind point spread function estimation," in *2019 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2019, pp. 2871–2875. (Cited on page(s): xii, 52, 84, 87, 89)

[C2] J. Hatvani, A. Basarab, J. Michetti, M. Gyöngy, and D. Kouamé, "Single image super-resolution of noisy 3D dental CT images using tucker decomposition," *IEEE International Synopsium on Biomedical Imaging (ISBI) [accepted]*, 2021. (Cited on page(s): xii, 52, 71, 85, 87, 89)

Other publications of the author

[Au1] J. Hatvani, "Applicability of computed tomography methods for tomosynthesis problems," in *PhD Proceedings Annual Issues of the Doctoral School, Faculty of Information Technology and Bionics, Pázmány Péter Catholic University – 2017*, 2017, pp. 48–48. (Cited on page(s): xii, 10)

[Au2] J. Hatvani, "Iterative optimization techniques for limited angle computed tomography," *Phd Proceedings Annual Issues of the Doctoral School Faculty of Information Technology and Bionics, Pázmány Péter Catholic University – 2018*, vol. 13, p. 45, 2018. (Cited on page(s): xii, 10)

[Au3] J. Hatvani, "A review of super-resolution techniques on computed tomography images," *Phd Proceedings Annual Issues of the Doctoral School Faculty of Information Technology and Bionics, Pázmány Péter Catholic University – 2019*, p. 46, 2019. (Cited on page(s): xii, 14)

[Au4] J. Hatvani, "A summary of tensor factorisation techniques for 3D image processing," in *PhD Proceedings Annual Issues of the Doctoral School, Faculty of Information Technology and Bionics, Pázmány Péter Catholic University – 2020 In press*, 2020. (Cited on page(s): xii, 27)

Bibliography

[1] T. L. Szabo, *Diagnostic ultrasound imaging: inside out.* Academic Press, 2004. (Cited on page(s): 1, 2)

[2] W. W. Moses, "Fundamental limits of spatial resolution in PET," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 648, pp. S236–S240, 2011. (Cited on page(s): 2)

[3] OECD. (2020) Computed tomography (CT) exams (indicator). Accessed: 2020. 05. 23. [Online]. Available: doi.org/10.1787/3c994537-en (Cited on page(s): 5, 6)

[4] D. J. Brenner and E. J. Hall, "Computed tomography—an increasing source of radiation exposure," *New England Journal of Medicine*, vol. 357, no. 22, pp. 2277–2284, 2007. (Cited on page(s): 6)

[5] M. S. Pearce, J. A. Salotti, M. P. Little, K. McHugh, C. Lee, K. P. Kim, N. L. Howe, C. M. Ronckers, P. Rajaraman, A. W. Craft *et al.*, "Radiation exposure from CT scans in childhood and subsequent risk of leukaemia and brain tumours: a retrospective cohort study," *The Lancet*, vol. 380, no. 9840, pp. 499–505, 2012. (Cited on page(s): 6)

[6] R. Behling, *Modern Diagnostic X-Ray Sources: Technology, Manufacturing, Reliability.* CRC Press, 2015. (Cited on page(s): 7)

[7] J. T. Bushberg and J. M. Boone, *The essential physics of Medical Imaging.* Lippincott Williams & Wilkins, 2011. (Cited on page(s): 7)

[8] J. Hsieh, *Computed tomography: principles, design, artifacts, and recent advances*. SPIE press, 2003, vol. 114. (Cited on page(s): 8)

[9] W. van Aarle, K. J. Batenburg, G. Van Gompel, E. d. Van Casteele, and J. Sijbers, "Super-resolution for computed tomography based on discrete tomography," *IEEE Transactions on Image Processing*, vol. 23, no. 3, pp. 1181–1193, March 2014. (Cited on page(s): 8, 10)

[10] J. Frikel, "Reconstructions in limited angle x-ray tomography: Characterization of classical reconstructions and adapted curvelet sparse regularization," Ph.D. dissertation, Technische Universität München, Fakultät für Mathematik, 2012. (Cited on page(s): 9)

[11] Q. B. Carroll, *Radiography in the Digital Age: Physics-exposure-radiation biology*. Charles C Thomas Publisher, 2018. (Cited on page(s): 11)

[12] L. W. Goldman, "Principles of ct: radiation dose and image quality," *Journal of nuclear medicine technology*, vol. 35, no. 4, pp. 213–225, 2007. (Cited on page(s): 11)

[13] M. F. McNitt-Gray, "Aapm/rsna physics tutorial for residents: topics in ct: radiation dose in ct," *Radiographics*, vol. 22, no. 6, pp. 1541–1553, 2002. (Cited on page(s): 11)

[14] A. Staude and J. Goebbels, "Determining the spatial resolution in computed tomography-comparison of mtf and line-pair structures," in *International Symposium on Digital Industrial Radiology and Computed Tomography - Tu.4.1*, 2011, pp. 1–9. (Cited on page(s): 12, 13)

[15] O. A. Peters, "Current challenges and concepts in the preparation of root canal systems: a review," *Journal of endodontics*, vol. 30, no. 8, pp. 559–567, 2004. (Cited on page(s): 13)

[16] Y.-L. Ng, V. Mann, S. Rahbaran, J. Lewsey, and K. Gulabivala, "Outcome of primary root canal treatment: systematic review of the literature–part 1. ef-

fects of study characteristics on probability of success," *International endodontic journal*, vol. 40, no. 12, pp. 921–939, 2007. (Cited on page(s): 13)

[17] H. Eriksen M, L. Kirkevang, and K. Petersson, "Endodontic epidemiology and treatment outcome: general considerations," *Endodontic Topics*, vol. 2, no. 1, pp. 1–9, 2002. (Cited on page(s): 13)

[18] E. Cotti, C. Dessì, A. Piras, and G. Mercuro, "Can a chronic dental infection be considered a cause of cardiovascular disease? a review of the literature," *International journal of cardiology*, vol. 148, no. 1, pp. 4–10, 2011. (Cited on page(s): 13)

[19] E. Cotti, C. Dessì, A. Piras, G. Flore, M. Deidda, C. Madeddu, A. Zedda, G. Longu, and G. Mercuro, "Association of endodontic infection with detection of an initial lesion to the cardiovascular system," *Journal of endodontics*, vol. 37, no. 12, pp. 1624–1629, 2011. (Cited on page(s): 13)

[20] J. J. Segura-Egea, L. Castellanos-Cosano, G. Machuca, J. López-López, J. Martín-González, E. Velasco-Ortega, B. Sánchez-Domínguez, and F. J. López-Frías, "Diabetes mellitus, periapical inflammation and endodontic treatment outcome," *Medicina oral, patología oral y cirugía bucal*, vol. 17, no. 2, p. e356, 2012. (Cited on page(s): 13)

[21] R. D. Astolphi, M. M. Curbete, N. H. Colombo, D. J. Shirakashi, F. Y. Chiba, A. K. C. Prieto, L. T. A. Cintra, S. R. M. Bomfim, E. Ervolino, and D. H. Sumida, "Periapical lesions decrease insulin signal and cause insulin resistance," *Journal of endodontics*, vol. 39, no. 5, pp. 648–652, 2013. (Cited on page(s): 13)

[22] M. S. Gomes, T. C. Blattner, M. Sant'Ana Filho, F. S. Grecca, F. N. Hugo, A. F. Fouad, and M. A. Reynolds, "Can apical periodontitis modify systemic levels of inflammatory markers? a systematic review and meta-analysis," *Journal of endodontics*, vol. 39, no. 10, pp. 1205–1217, 2013. (Cited on page(s): 13)

[23] D. Brüllmann and R. K. W. Schulze, "Spatial resolution in CBCT machines for dental/maxillofacial applications —- what do we know today?" *Dentomaxillofacial Radiology*, vol. 44, no. 1, p. 20140204, 2014. (Cited on page(s): 13)

[24] J. Martos, G. H. Tatsch, A. C. Tatsch, L. F. M. Silveira, and C. M. Ferrer-Luque, "Anatomical evaluation of the root canal diameter and root thickness on the apical third of mesial roots of molars," *Anat. Sci. Int.*, vol. 86, no. 3, pp. 146–150, Sep. 2011. (Cited on page(s): 13)

[25] K. Horner and S. Panel, *Cone Beam CT for Dental and Maxillofacial Radiology (Evidence Based Guidelines)*, ser. Radiation Protection series. European Commission: Directorate-General for Energy, 2012, the Seventh Framework Programme of the European Atomic Energy Community (Euratom) for nuclear research and training activities (2007 to 2011) http://cordis.europa.eu/fp7/euratom/. (Cited on page(s): 13)

[26] T. Pan, T.-Y. Lee, E. Rietzel, and G. T. Chen, "4d-ct imaging of a volume influenced by respiratory motion on multi-slice ct," *Medical physics*, vol. 31, no. 2, pp. 333–340, 2004. (Cited on page(s): 14)

[27] C. H. McCollough, S. Leng, L. Yu, and J. G. Fletcher, "Dual-and multi-energy ct: principles, technical approaches, and clinical applications," *Radiology*, vol. 276, no. 3, pp. 637–653, 2015. (Cited on page(s): 14)

[28] K. I. Kim and Y. Kwon, "Single-image super-resolution using sparse regression and natural image prior," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 6, pp. 1127–1133, 2010. (Cited on page(s): 14)

[29] F. Shi, J. Cheng, L. Wang, P. Yap, and D. Shen, "LRTV: MR image super-resolution with low-rank and total variation regularizations," *IEEE Trans. Med. Imag.*, vol. 34, no. 12, pp. 2459–2466, 2015. (Cited on page(s): 14, 16, 17, 18, 32, 54)

[30] N. Zhao, Q. Wei, A. Basarab, N. Dobigeon, D. Kouamé, and J.-Y. Tourneret, "Fast single image super-resolution using a new analytical solution for ℓ_2–ℓ_2

problems," *IEEE Trans. Image Process.*, vol. 25, no. 8, pp. 3683–3697, Aug. 2016. (Cited on page(s): 14, 16, 17)

[31] X. Li and M. T. Orchard, "New edge-directed interpolation," *IEEE transactions on image processing*, vol. 10, no. 10, pp. 1521–1527, 2001. (Cited on page(s): 16)

[32] A. Toma, L. Denis, B. Sixou, J. Pialat, and F. Peyrin, "Total variation super-resolution for 3D trabecular bone micro-structure segmentation," in *Proc. 22nd European Signal Processing Conference (EUSIPCO)*, Lisbon, Portugal, Sep. 2014, pp. 2220–2224. (Cited on page(s): 17, 38)

[33] W. Zhang, H. Zhang, L. Wang, A. Cai, L. Li, and B. Yan, "Limited angle CT reconstruction by simultaneous spatial and radon domain regularization based on tv and data-driven tight frame," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 880, pp. 107–117, 2018. (Cited on page(s): 17)

[34] J. V. Manjón, P. Coupé, A. Buades, V. Fonov, D. L. Collins, and M. Robles, "Non-local MRI upsampling," *Medical image analysis*, vol. 14, no. 6, pp. 784–792, 2010. (Cited on page(s): 17)

[35] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on information theory*, vol. 52, no. 2, pp. 489–509, 2006. (Cited on page(s): 17)

[36] S. Sarraf and G. Tofighi, "Classification of alzheimer's disease using fMRI data and deep learning convolutional neural networks," *arXiv preprint arXiv:1603.08631*, 2016. (Cited on page(s): 19, 20)

[37] L. Zhao and K. Jia, "Multiscale CNNs for brain tumor segmentation and diagnosis," *Comput Math Methods Med.*, vol. 2016, 2016. (Cited on page(s): 19, 20)

[38] H. Choi and K. H. Jin, "Fast and robust segmentation of the striatum using deep convolutional neural networks," *J. Neurosci. Methods*, vol. 274, pp. 146–153, Dec. 2016. (Cited on page(s): 19, 20)

[39] M. Ghafoorian, N. Karssemeijer, T. Heskes, M. Bergkamp, J. Wissink, J. Obels, K. Keizer, F. E. Leeuw, B. V. Ginneken, E. Marchiori, and B. Platel, "Deep multi-scale location-aware 3D convolutional neural networks for automated detection of lacunes of presumed vascular origin," *NeuroImage: Clinical*, vol. 14, pp. 391–399, 2017. (Cited on page(s): 19, 20)

[40] P. Yuehao, H. Weimin, L. Zhiping, Z. Wanzheng, Z. Jiayin, J. Wong, and D. Zhongxiang, "Brain tumor grading based on neural networks and convolutional neural networks," in *Proc. 37th Annu. Int. IEEE EMBC Conf.*, Milano, Italy, Aug.25-29 2015, pp. 699–702. (Cited on page(s): 19, 20)

[41] A. Benou, R. Veksler, A. Friedman, and T. R. Raviv, "De-noising of contrast-enhanced MRI sequences by an ensemble of expert deep neural networks," in *Deep Learning and Data Labeling for Medical Applications: Proc. 1st Int. Workshop LABELS'16, and 2nd Int. Workshop DLMIA'16 Held in Conjunction with MICCAI'16*, Athens, Greece, Oct.21 2016, pp. 95–110. (Cited on page(s): 19)

[42] D. Nie, X. Cao, Y. Gao, L-Wang, and D. Shen, "Estimating CT image from MRI data using 3D fully convolutional networks," in *Proc. 1st Int. Workshop LABELS'16, and 2nd Int. Workshop DLMIA'16 Held in Conjunction with MICCAI'16*. Athens, Greece: Springer, Oct.21 2016, pp. 170–178. (Cited on page(s): 19, 20)

[43] K. H. Jin, M. T. McCann, E. Froustey, and M. Unser, "Deep convolutional neural network for inverse problems in imaging," *IEEE Trans. Image Process.*, vol. 26, no. 9, pp. 4509–4522, Sep. 2017. (Cited on page(s): 19)

[44] Y. Han and J. C. Ye, "Framing u-net via deep convolutional framelets: Application to sparse-view ct," *arXiv preprint arXiv:1708.08333*, 2017. (Cited on page(s): 19, 26)

[45] H. Zhang, L. Li, K. Qiao, L. Wang, B. Yan, L. Li, and G. Hu, "Image prediction for limited-angle tomography via deep learning with convolutional neural network," *arXiv preprint arXiv:1607.08707*, 2016. (Cited on page(s): 20)

[46] O. Oktay, W. Bai, M. Lee, R. Guerrero, K. Kamnitsas, J. Caballero, A. de Marvao, S. Cook, D. O'Regan, and D. Rueckert, "Multi-input cardiac image super-resolution using convolutional neural networks," in *Proc. 19th MICCAI Int. Conf.* Athens, Greece: Springer, Oct.17–21 2016, pp. 246–254. (Cited on page(s): 20)

[47] S. Cengiz, M. d. C. Valdes-Hernandez, and E. Ozturk-Isik, "Super resolution convolutional neural networks for increasing spatial resolution of ^{1}h magnetic resonance spectroscopic imaging," in *Proc. 21st MIUA Annu. Conf.* Edinburgh, UK: Springer, Jul.11-13 2017, pp. 641–650. (Cited on page(s): 20)

[48] Y. Zhang and M. An, "Deep learning- and transfer learning-based super resolution reconstruction from single medical image," *J. Healthc. Eng.*, vol. 2017, Art. ID 5859727, 20 pages, 2017. (Cited on page(s): 20)

[49] B. Xu, N. Wang, T. Chen, and M. Li, "Empirical evaluation of rectified activations in convolutional network," *arXiv preprint arXiv:1505.00853*, 2015. (Cited on page(s): 21)

[50] H. Zhao, O. Gallo, T. Frosio, and J. Kautz, "Loss functions for neural networks for image processing," *arXiv preprint arXiv:1511.08861*, 2015. (Cited on page(s): 22)

[51] P. J. Huber, "The place of the l1-norm in robust estimation," *Computational Statistics & Data Analysis*, vol. 5, no. 4, pp. 255–262, Sep. 1987. (Cited on page(s): 22)

[52] D. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014. (Cited on page(s): 23)

[53] Y. Li, B. Sixou, and F. Peyrin, "A review of the deep learning methods for medical images super resolution problems," *IRBM*, 2020. [Online]. Available:

https://www.sciencedirect.com/science/article/pii/S1959031820301408 (Cited on page(s): 25, 26)

[54] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 2, pp. 295–307, Feb. 2016. (Cited on page(s): 25, 80, 83)

[55] J. Kim, J. K. Lee, and K. M. Lee, "Accurate image super-resolution using very deep convolutional networks," in *Proc. 29th IEEE CVPR Conf.*, Las Vegas, US, Jun. 2016, pp. 1646–1654. (Cited on page(s): 25)

[56] ——, "Deeply-recursive convolutional network for image super-resolution," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 1637–1645. (Cited on page(s): 25)

[57] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in neural information processing systems*, vol. 25, pp. 1097–1105, 2012. (Cited on page(s): 26)

[58] W. Shi, J. Caballero, F. Huszár, J. Totz, A. P. Aitken, R. Bishop, D. Rueckert, and Z. Wang, "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network," in *Proc. 29th IEEE CVPR Conf.*, Las Vegas, US, Jun. 2016, pp. 1874–1883. (Cited on page(s): 26, 42, 83)

[59] W. Shi, J. Caballero, L. Theis, F. Huszar, A. Aitken, C. Ledig, and Z. Wang, "Is the deconvolution layer the same as a convolutional layer?" *arXiv preprint arXiv:1609.07009*, 2016. (Cited on page(s): 26, 42, 43)

[60] T. Tong, G. Li, X. Liu, and Q. Gao, "Image super-resolution using dense skip connections," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 4799–4807. (Cited on page(s): 26)

[61] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proc. 18th Int. Conf. MICCAI*. Munich, Germany: Springer, Oct.5–9 2015, pp. 234–241. (Cited on page(s): 26, 41, 42, 83)

[62] C. I. Kanatsoulis, X. Fu, N. D. Sidiropoulos, and W. Ma, "Hyperspectral super-resolution: A coupled tensor factorization approach," *IEEE Trans. Signal Process.*, vol. 66, no. 24, pp. 6503–6517, Dec 2018. (Cited on page(s): 27, 32, 54, 84)

[63] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM review*, vol. 51, no. 3, pp. 455–500, 2009. (Cited on page(s): 27, 29, 32, 72)

[64] J. D. Carroll and J.-J. Chang, "Analysis of individual differences in multidimensional scaling via an n-way generalization of "eckart-young" decomposition," *Psychometrika*, vol. 35, no. 3, pp. 283–319, 1970. (Cited on page(s): 31, 32)

[65] L. Chiantini and G. Ottaviani, "On generic identifiability of 3-tensors of small rank," *SIAM Journal on Matrix Analysis and Applications*, vol. 33, no. 3, pp. 1018–1037, 2012. (Cited on page(s): 31)

[66] R. Bro, C. A. Andersson, and H. A. Kiers, "Parafac2—part ii. modeling chromatographic data with retention time shifts," *Journal of Chemometrics: A Journal of the Chemometrics Society*, vol. 13, no. 3-4, pp. 295–309, 1999. (Cited on page(s): 32)

[67] A. H. Andersen and W. S. Rayens, "Structure-seeking multilinear methods for the analysis of fmri data," *NeuroImage*, vol. 22, no. 2, pp. 728–739, 2004. (Cited on page(s): 32)

[68] E. Martınez-Montes, P. A. Valdés-Sosa, F. Miwakeichi, R. I. Goldman, and M. S. Cohen, "Concurrent eeg/fmri analysis by multiway partial least squares," *NeuroImage*, vol. 22, no. 3, pp. 1023–1034, 2004. (Cited on page(s): 32)

[69] A. Shashua and A. Levin, "Linear image coding for regression and classification using the tensor-rank principle," in *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, vol. 1. IEEE, 2001, pp. I–I. (Cited on page(s): 32)

[70] L. R. Tucker, "Some mathematical notes on three-mode factor analysis," *Psychometrika*, vol. 31, no. 3, pp. 279–311, 1966. (Cited on page(s): 33)

[71] B. W. Bader and T. G. Kolda, "Efficient matlab computations with sparse and factored tensors," *SIAM Journal on Scientific Computing*, vol. 30, no. 1, pp. 205–231, 2008. (Cited on page(s): 33)

[72] M. A. O. Vasilescu and D. Terzopoulos, "Multilinear analysis of image ensembles: Tensorfaces," in *European conference on computer vision*. Springer, 2002, pp. 447–460. (Cited on page(s): 33)

[73] J. Liang, Y. He, D. Liu, and X. Zeng, "Image fusion using higher order singular value decomposition," *IEEE transactions on Image Processing*, vol. 21, no. 5, pp. 2898–2909, 2012. (Cited on page(s): 34)

[74] A. Rövid, I. J. Rudas, S. Sergyán, and L. Szeidl, "Hosvd based image processing techniques," in *Proc. of the 10th WSEAS International Conference on Artificial Intelligence, Knowledge Engineering and Data Bases*, 2011, pp. 297–302. (Cited on page(s): 34)

[75] Y. Zhang, H. Chan, B. Sahiner, J. Wei, M. Goodsitt, L. Hadjiiski, J. Ge, and C. Zhou, "A comparative study of limited-angle cone-beam reconstruction methods for breast tomosynthesis," *Med Phys*, vol. 33, no. 10, pp. 3781–3795, Oct 2006. (Cited on page(s): 34)

[76] C. Prévost, K. Usevich, P. Comon, and D. Brie, "Hyperspectral super-resolution with coupled tucker approximation: Recoverability and svd-based algorithms," *IEEE Transactions on Signal Processing*, vol. 68, pp. 931–946, 2020. (Cited on page(s): 34)

[77] R. Kikinis, S. D. Pieper, and K. G. Vosburgh, "3D slicer: a platform for subject-specific image analysis, visualization, and clinical support," in *Intraoperative imaging and image-guided therapy*. Springer New York, 2014, ch. 19, pp. 277–289. (Cited on page(s): 36)

[78] A. Fedorov, R. Beichel, J. Kalpathy-Cramer, J. Finet, J. C. Fillion-Robin, S. Pujol, C. Bauer, D. Jennings, F. Fennessy, M. Sonka *et al.*, "3D slicer as an

image computing platform for the quantitative imaging network," *Magnetic resonance imaging*, vol. 30, no. 9, pp. 1323–1341, Nov. 2012. (Cited on page(s): 36)

[79] (2017) 3D slicer. [Online]. Available: https://www.slicer.org/ (Cited on page(s): 36)

[80] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004. (Cited on page(s): 39)

[81] H. R. Sheikh, A. C. Bovik, and G. D. Veciana, "An information fidelity criterion for image quality assessment using natural scene statistics," *IEEE Trans. Image Process.*, vol. 14, no. 12, pp. 2117–2128, Dec. 2005. (Cited on page(s): 39)

[82] N. Damera-Venkata, T. D. Kite, W. S. Geisler, B. L. Evans, and A. C. Bovik, "Image quality assessment based on a degradation model," *IEEE Trans. Image Process.*, vol. 9, no. 4, pp. 636–650, Apr. 2000. (Cited on page(s): 40)

[83] J. Michetti, A. Basarab, F. Diemer, and D. Kouame, "Comparison of an adaptive local thresholding method on CBCT and μCT endodontic images," *Physics in Medicine & Biology*, vol. 63, no. 1, p. 015020, 2017. (Cited on page(s): 40)

[84] L. Dice, "Measures of the amount of ecologic association between species," *Ecology*, vol. 26, no. 3, pp. 297–302, 1945. (Cited on page(s): 40)

[85] MeVis Medical Solutions AG. (2017) MeVisLab. Available online. [Online]. Available: https://www.mevislab.de (Cited on page(s): 40)

[86] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. 32nd Int. Conf. on Machine Learning ICML*, Lille, France, 2015, pp. 448–456. (Cited on page(s): 41, 44)

[87] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 2818–2826. (Cited on page(s): 41)

[88] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proc. 13th Int. Conf. on Artificial Intelligence and Statistics*, 2010, pp. 249–256. (Cited on page(s): 43)

[89] M. Abadi, A. Agarwal, P. Barham *et al.*, "TensorFlow: Large-scale machine learning on heterogeneous systems," 2015, software available from tensorflow.org. [Online]. Available: https://www.tensorflow.org/ (Cited on page(s): 44)

[90] J. H. Elder and S. W. Zucker, "Local scale control for edge detection and blur estimation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 7, pp. 699–716, 1998. (Cited on page(s): 53)

[91] M. Fuhry and L. Reichel, "A new tikhonov regularization method," *Numerical Algorithms*, vol. 59, no. 3, pp. 433–445, 2012. (Cited on page(s): 55)

[92] N. Vervliet, O. Debals, L. Sorber, M. Van Barel, and L. De Lathauwer. (2016, Mar.) Tensorlab 3.0. Available online. [Online]. Available: https://www.tensorlab.net (Cited on page(s): 57, 68)

[93] N. Zhao, Q. Wei, A. Basarab, D. Kouamé, and J.-Y. Tourneret, "Blind deconvolution of medical ultrasound images using a parametric model for the point spread function," in *Ultrasonics Symposium (IUS), 2016 IEEE International*. IEEE, 2016, pp. 1–4. (Cited on page(s): 66)

[94] C. Tsutake and T. Yoshida, "Vaguelette-wavelet deconvolution via compressive sampling," *IEEE Access*, vol. 7, pp. 54 533–54 541, 2019. (Cited on page(s): 71)

[95] R. James, A. M. Jolly, A. C., and D. Michael, "Image denoising using adaptive pca and svd," in *2015 Fifth International Conference on Advances in Computing and Communications (ICACC)*, Sep. 2015, pp. 383–386. (Cited on page(s): 72)

[96] M. Holmes, A. Gray, and C. Isbell, "Fast svd for large-scale matrices," in *Workshop on Efficient Machine Learning at NIPS*, vol. 58, 2007, pp. 249–252. (Cited on page(s): 72)

[97] Z. Drmac and K. Veselic, "New fast and accurate jacobi svd algorithm. i," *SIAM Journal on matrix analysis and applications*, vol. 29, no. 4, pp. 1322–1342, 2008. (Cited on page(s): 72)

[98] A. K. Menon and C. Elkan, "Fast algorithms for approximating the singular value decomposition," *ACM Trans. Knowl. Discov. Data*, vol. 5, no. 2, Feb. 2011. [Online]. Available: https://doi.org/10.1145/1921632.1921639 (Cited on page(s): 75)

[99] J. A. Grant-Jacob, B. S. Mackay, J. A. Baker, Y. Xie, D. J. Heath, M. Loxham, R. W. Eason, and B. Mills, "A neural lens for super-resolution biological imaging," *Journal of Physics Communications*, vol. 3, no. 6, p. 065004, 2019. (Cited on page(s): 78)

[100] Z. Zhou, Y. Wang, Y. Guo, Y. Qi, and J. Yu, "Image quality improvement of hand-held ultrasound devices with a two-stage generative adversarial network," *IEEE Transactions on Biomedical Engineering*, vol. 67, no. 1, pp. 298–311, 2019. (Cited on page(s): 78)

[101] M.-I. Georgescu, R. T. Ionescu, and N. Verga, "Convolutional neural networks with intermediate loss for 3d super-resolution of ct and mri scans," *IEEE Access*, vol. 8, pp. 49 112–49 124, 2020. (Cited on page(s): 78)

[102] J. Michetti, A. Basarab, M. Tran, F. Diemer, and D. Kouamé, "Cone-beam computed tomography contrast validation of an artificial periodontal phantom for use in endodontics," in *Proc. IEEE Engineering in Medicine and Biology Society (EMBC), 37th Annual Int. Conf.* IEEE, 2015, pp. 7905–7908. (Cited on page(s): 80)

[103] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang *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, pp. 4681–4690. (Cited on page(s): 80)

[104] F. Bazzi, M. Mescam, A. Basarab, and D. Kouamé, "On single-image super-resolution in 3D brain magnetic resonance imaging," in *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2019, pp. 2840–2843. (Cited on page(s): 81)

Résumé - La résolution spatiale des images acquises par tomographie volumique á faisceau conique (CBCT) est limitée par la géométrie des capteurs, leur sensibilité, les mouvements du patient, les techniques de reconstruction d'images et la limitation de la dose de rayonnement. Le modéle de dégradation d'image considéré dans cette these consiste en un opérateur de ou avec la fonction d'étalement du systéme d'imagerie (PSF), un opérateur de décimation, et du bruit, qui relient les volumes CBCT á une image 3D super-résolue á estimer. Les méthodes proposées dans cette thése (SISR - single image super-resolution) ont comme objectif d'inverser ce modéle direct, c'est á dire d'estimer un volume haute résolution á partir d'une image CBCT. Les algorithms ont été évalués dans le cadre d'une application dentaire, avec comme verité terrain les images haute résolution acquises par micro CT (μCT), qui utilise des doses de rayonnement trés importantes, incompatibles avec les applications cliniques.

Mots clés: tomodensitométrie, CBCT dentaire, image volumétrique, super-résolution, factorization tensorielle, apprentissage profond

Abstract - The resolution of dental cone beam computed tomography (CBCT) images is limited by detector geometry, sensitivity, patient movement, the reconstruction technique and the need to minimize radiation dose. The corresponding image degradation model assumes that the CBCT image is a blurred (with a point spread function, PSF), downsampled, noisy version of a high resolution image. The quality of the image is crucial for precise diagnosis and treatment planning. The methods proposed in this thesis aim to give a solution for the single image super-resolution (SISR) problem. The algorithms were evaluated on dental CBCT and corresponding highresolution (and high radiation-dose) μCT image pairs of extracted teeth.

Keywords: Computed tomography, dental CBCT, 3D image, super-resolution, tensor factorization, deep learning

Institut de Recherche en Informatique de Toulouse,
118 Route de Narbonne, Toulouse