MEDICAL IMAGE ENHANCEMENT USING DEEP LEARNING AND TENSOR FACTORIZATION TECHNIQUES

Janka Hatvani

Theses of the Ph.D. Dissertation

Supervisors Dr. Miklós Gyöngy

Dr. Adrian Basarab

Prof. Denis Kouamé

Pázmány Péter Catholic University

Roska Tamás Doctoral School of Sciences and Technology

University of Toulouse III

Toulouse Doctoral School of Mathematics, Informatics and Telecommunication

Budapest, 2021

Introduction

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 are promoted.

Endodontics is a good example to show the importance of high quality images in medical treatments and diagnosis. In dentistry the 3D structure of the tooth is visualized using cone beam computed tomography (CBCT), where the typical resolution is around 500 µm [1]. 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 [2]. Even though endodontic treatment is one of the most common procedures, epidemiological studies show success rates of only 60-85% for general practice [3].

The resolution enhancement techniques presented in the thesis are demonstrated on dental CT scans, but each of them is adaptable to various modalities.

Challenges in Medical Image Enhancement

Recording higher quality images either requires expensive new devices (e.g. denser detector arrays), poses health risks to the patient (e.g. a higher dose of ionizing radiation), or is limited by physical boundaries (like the diffraction limit), therefore post-processing resolution enhancement is preferred.

The degradation model of the recorded images 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 [4]. State-of-the-art techniques are computationally efficient methods in the case of two-dimensional (2D) images. However, most medical images are three-dimensional (3D), and the size of the data volume does not permit the use of current SISR techniques in real life scenarios because of the extreme run-times (hours for a single dental volume [5]).

In the light of the above, the central research questions 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?

New Scientific Results

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 (CNN) have shown promising results for resolution enhancement [6]. To our knowledge this was the first time that a deep learning algorithm was used for biomedical SISR.

The U-net network [7] 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 [8] 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 2D reconstruction-based super-resolution methods (SRR) using ℓ_2-norm and total variation (TV) regularization were used for comparison. Some example outputs are shown in Fig. 1.

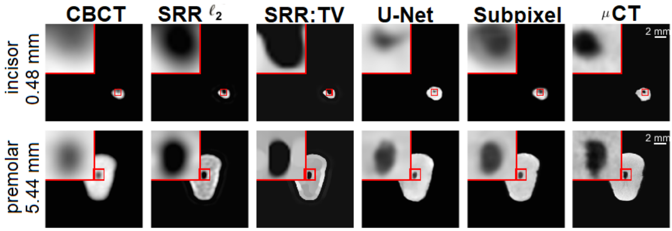

Figure 1: Result of SR methods on different slices from the test set. On the left of the first column the type of the tooth and the depth of the slice from the apex of the root are displayed. A 2 mm-scalebar is displayed on the μCT images. The display range is stretched to [0,1].

The results were evaluated using different metrics (in Table 1), in particular the peak signal-to-noise ratio (PSNR [dB]), structural similarity index (SSI) , and the difference in canal sizes (DoC) and Dice coefficient (DC) of subsequent 3D canal segmentation (Fig. 2).

Table 1: Quantitative DL enhancement results

Metrics averaged on the test set, compared to the µCT.

Metric	CBCT	SRR:ℓ_2	SRR:TV	U-net	Subpixel
PSNR	45.56	64.15	64.80	**67.58**	66.60
SSI	0.9145	0.8688	0.8830	0.9304	**0.9346**
DoC	12.39%	12.25%	12.40%	10.12%	**6.07%**
DC	0.8891	0.8852	0.8913	0.8998	**0.9101**

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

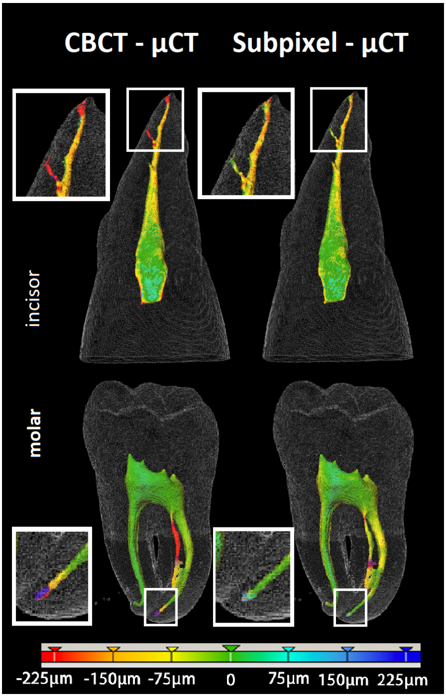

Figure 2: Volumetric segmentation of the root canal on two teeth of the test set. Coloring shows the difference between CBCT and μCT (on the left) and between the subpixel CNN and μCT segmentations.

medically salient features such as the size, shape, or curvature of the root canal, especially in the critical apical region. 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 hypetspectral images [9]. 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.

Table 2: Quantitative CPD-SISR results

	Sample #1	Sample #3
tooth type	upper incisor	lower molar
µCT image size	282×266×392	324×306×402
CBCT PSNR	23.17 dB	24.14 dB
LRTV PSNR	**24.32** dB	24.61 dB
CPD-SISR PSNR	**24.32** dB	**25.71** dB
CBCT DC	0.88 dB	0.90 dB
LRTV DC	0.87dB	0.90 dB
CPD-SISR DC	**0.90** dB	**0.91** dB
LRTV time	6988 s	10301 s
CPD-SISR time	**71 s**	**104 s**

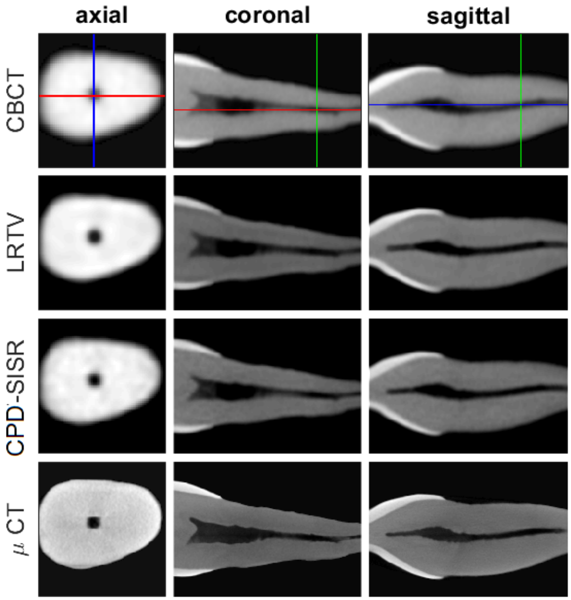

Figure 3: Results on a sample image (#1). The CBCT image is shown at the higher scale of the HR images, for better comparison. The location of the slices within the volume is illustrated on the CBCT images in colored lines.

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 (Fig. 3 and 4). 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 (Table 2. These results were promising enough for further research, as described in the following thesis points.

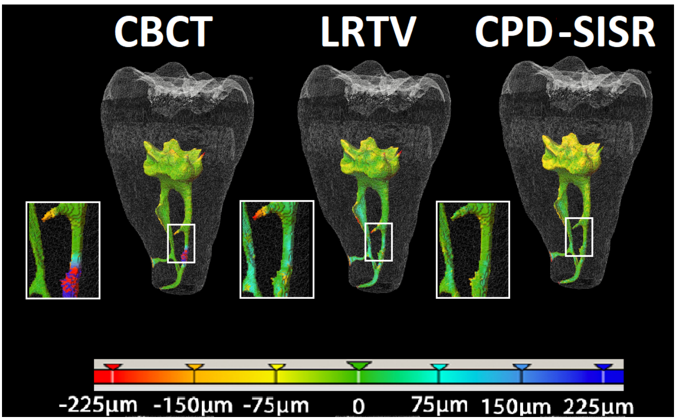

Figure 4: Segmentation results for CBCT, LRTV and CPD-SISR for a sample tooth (#3). The color-bar visualizes the distance between the estimated surface of the canal and the one obtained with μCT segmentation.

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.

Table 3: CPD-SISR-blind quantitative results

	Simulation	Experiment
µCT size	287×266×392	274×278×474
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**

In this work a semi-blind estimation was realized, assuming that the standard deviations of the Gaussian PSF ($\overline{\sigma}$) are within a known interval. The problem optimizing for these parameters can be solved with gradient descent [10]. This minimization for the PSF and the CPD-SISR optimizing for the high-resolution image are repeated alternating until convergence.

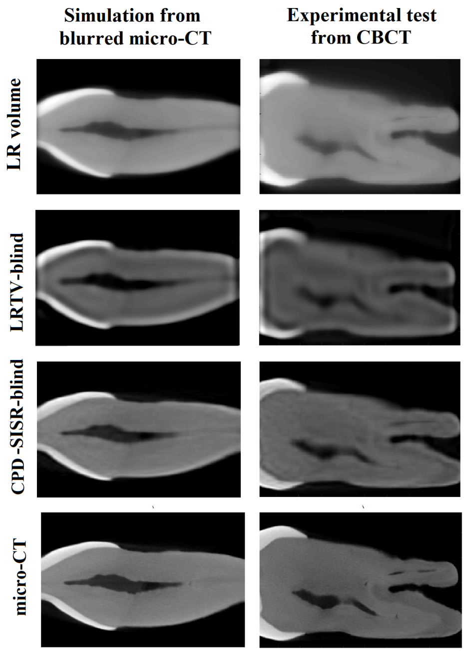

Figure 5: Qualitative results with CPD-SISR-blind. A coronal slice was chosen for demonstration. The LR images (artificially degraded μCT in simulation, CBCT in experimental test) are shown on the scale of the μCT images using linear interpolation.

For comparison the LRTV algorithm was used as in Thesis II a, combined with the proposed PSF-estimation in a similar alternating manner, noted as LRTV-blind (Fig. 5). 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 (Table 3). 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 [11]. The basis vectors may be weighted according to their importance in the factorization. While the CPD defines a single rank that has to be estimated, TD uses the nrank, three different values for 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, therefor they are implemented subsequently.

Table 4: Quantitative results in TD-SISR - 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**

Even though two additional parameters have to be set, it gave faster and quantitatively better results in noisy simulated and real images compared to the previous method, CPD-SISR (Table 5, Fig. 6). Images of 280×268×492 and 324×248×442 pixels were super-resolved under 2 s with standard Matlab implementation. The PSNR has improved under added noise in both methods. TD-SISR outperformed CPD-SISR both in PSNR and SSI values, except for the extremely noisy, 20 dB case. The segmentation was carried out at 25 dB. The improvement is confirmed by the DC, showing the superiority of the TD-SISR method.

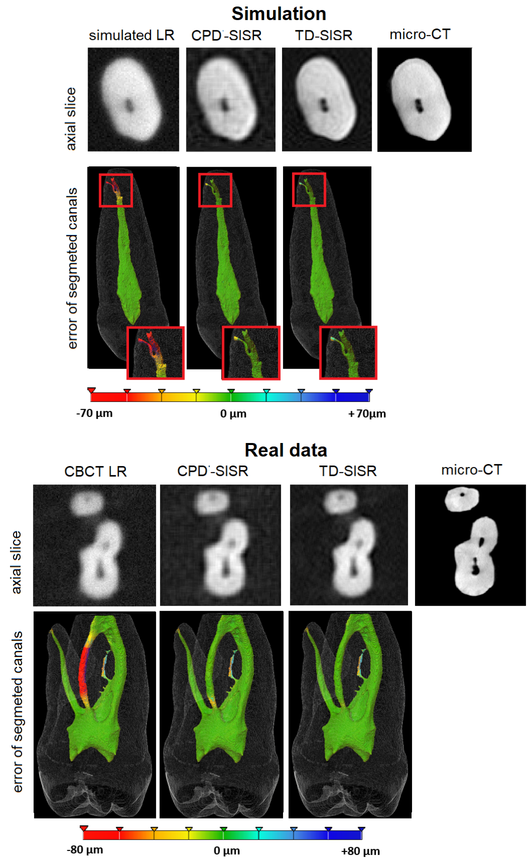

Figure 6: Results of SISR methods under 25 dB noise, both in simulation and in real data. The first row shows a single axial slice taken from the volumes. The second row shows the distance between the segmented high-resolution, low-resolution, enhanced volumes.

Table 5: Quantitative results in TD-SISR - real data

	CBCTR	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	
DC	0.8939	0.9189	**0.9304**

Publications related to 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.

[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.

[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.

[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," *arXiv preprint arXiv:2009.08657-eess.IV*, 2020.

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.

[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.

[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.

[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.

References

[1] 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.

[2] 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.

[3] 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. effects of study characteristics on probability of success," *International endodontic journal*, vol. 40, no. 12, pp. 921–939, 2007.

[4] 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.

[5] 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.

[6] 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.

[7] 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.

[8] 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.

[9] 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.

[10] 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.

[11] L. R. Tucker, "Some mathematical notes on three-mode factor analysis," *Psychometrika*, vol. 31, no. 3, pp. 279–311, 1966.