Deep Learning-Based Super-Resolution Applied to Dental Computed Tomography

Janka Hatvani, András Horváth, Jérôme Michetti, Adrian Basarab, Denis Kouamé, and Miklós Gyöngy

***Abstract*—The resolution of dental computed tomography (CT) images is limited by detector geometry, sensitivity, patient movement, the reconstruction technique and the need to minimize radiation dose. Recently, the use of convolutional neural network (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 2-D cone-beam CT 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 total variation 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 image-segmentation-based analysis. In the evaluation, micro-CT 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 dental CT images, allowing better detection of medically salient features, such as the size, shape, or curvature of the root canal.**

***Index Terms*—Computed tomography (CT), convolutional neural networks (CNNs), deconvolution, dental applications, image analysis, super-resolution (SR), U-net.**

I. Introduction

ENDODONTICS is the dental specialty concerned with the maintenance of the dental pulp (formed by nerves, blood vessels, and connective tissues) in healthy state and with the treatment of the pulp cavity, i.e., pulp chamber and root canal (the internal part of the tooth). A good knowledge of the root canal anatomy is an indispensable prerequisite for ensuring the success of pulp cavity treatment. According to [1], three guidelines are important and must be followed during such treatment: 1) identifying and preparing the root main canals using endodontic instruments; 2) establishing and respecting working length; and 3) assessing the initial apical canal diameter to allow an adequate preparation size. Even though endodontic treatment is one of the most common procedures, epidemiological studies show success rates of only 60%–85% for general practice [2], [3]. The reduction of endodontic therapeutic failures, i.e., 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, require new techniques for improving the quality of endodontic treatments [4]–[8]. In dentistry the 3-D structure of the tooth is visualized using cone beam computed tomography (CBCT), where the typical resolution is around 500 μm because of partial volume effect, noise, and beam hardening [9]. 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 [10]. 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 [11]. Micro-CT (μCT) gives a sufficient resolution for precise segmentation of the pulp cavity but can be used only *ex vivo* (on extracted teeth) due to size limitations, long acquisition time, and high radiation dose. An algorithmic solution for approximating the resolution of μCT images from CBCT acquisitions would therefore be advantageous.

Super resolution (SR)—finding the high resolution (HR) image from a single or multiple low resolution (LR) image(s)—is a well-known problem in image processing. The main groups of methods use interpolation with edge preservation, deconvolution-based reconstruction with Bayesian predictions or regularization, and example- or patch-dictionaries [12]. In the last group the LR and HR patches are mapped nonlinearly onto each other after feature extraction, reconstructing the final solution from the HR patches. Until recent years, the state-of-the-art solution for the problem used this approach with sparsity-based machine learning, despite its high computational complexity and its dependency on the training set [13].

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,

Manuscript received November 14, 2017; revised February 23, 2018 and March 27, 2018; accepted April 6, 2018. Date of publication April 16, 2018; date of current version March 1, 2019. This work was supported in part by the Hungarian and EU Funding (Széchenyi 2020 Program) under Grant EFOP-3.6.2-16-2017-00013 and Grant 3.6.3-VEKOP-16-2017-00002, in part by the Pázmány University under Grant KAP16-18, and in part by the Thematic Trimester on Image Processing of the CIMI Labex, Toulouse, France, with the Program ANR-11-IDEX-0002-02 under Grant ANR-11-LABX-0040-CIMI. The work of A. Horváth and M. Gyöngy was supported in part by the Bolyai Scholarship of the Hungarian Academy of Sciences under Grant PD 121105, and in part by the Hungarian National Research, Development and Innovation Office. *(Corresponding author: Janka Hatvani.)*

J. Hatvani, A. Horváth, and M. Gyöngy are with the Faculty of Information Technology and Bionics, Pazmany Peter Catholic Unviersity, 1088 Budapest, Hungary (e-mail: hatvani.janka@itk.ppke.hu).

J. Michetti, A. Basarab, and D. Kouamé are with the Institut de Recherche en Informatique de Toulouse, University Paul Sabatier, 31062 Toulouse, France.

Digital Object Identifier 10.1109/TRPMS.2018.2827239

opening a new perspective for SR techniques as well [14]. In biomedical imaging CNNs are mainly used for classification, segmentation and detection. Litjens *et al.* [15] gave a comprehensive overview on the topic. Some examples for these kinds of tasks are differential diagnosis between Alzheimer's and Huntington's diseases on MRI data [16], [17], tumor segmentation with multiscale analysis [18], striatum segmentation [19], or tumor and lesion detection [20], classification [21].

While deep learning is increasingly practised 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 [13], [22], image generation, e.g., constructing CT images from MRI data [23], or artifact removal from sparse-view [24], [25] and limited-angle CT images [26]. To the authors' knowledge, biomedical image resolution enhancement with deep learning has so far only been implemented using multi-input frameworks—the input was either an LR cardiac MRI sequence [27], [28] or multichannel MRI data (T1-, T2-weighted and fluid attenuated inversion recovery images [29]). Most of the previously mentioned biomedical applications tend to use CNNs [16], [18]–[21], [23], [27].

CNNs can realize the previously described SR pipeline of image-enhancement, namely feature extraction, nonlinear mapping, and reconstruction. In a CNN, the units within a layer are organized in such a way that the multiplication of input pixels with their corresponding weights implements a convolution process followed by a nonlinear activation operator, passing a series of filtered images to the upcoming layer. The output of the combined layers can either be an image or a classification answer. The weights of the convolution layers and the classification are learned via error backpropagation.

Many different approaches have been investigated for the enhancement of training in terms of quality and speed. In the pioneering, 2014 work of Dong *et al.* [14] the single image SR algorithm starts with the upsampling of the LR image using bicubic interpolation. However, Shi *et al.* [30] showed in 2016 that this step can be left out. Kim *et al.* [32] have shown that the number of layers can increase the performance, so that deep CNNs highly outperform the shallow ones [48]. An interesting structure called U-net was introduced for biomedical image segmentation and artifact removal, where features on different scales are learned efficiently [25], [33].

The aim of this paper was to investigate the use of CNNs for resolution enhancement of 2-D CBCT dental images, using μCT data of the same teeth as ground truth. Two different network structures have been tested, a subpixel network and a U-net designed for the given task. Its outputs were compared to deconvolution-based reconstruction techniques with ℓ_2-norm and total variation (TV) regularization.

In the rest of this paper, the methods are first described, namely the collection of data, the reconstruction-based SR methods used as reference, the tested deep learning architectures, and finally the metrics used for comparison. In the Results, the evolution of the network training is presented, followed by qualitative and quantitative comparisons, and segmentation-based image analysis.

II. Methods

A. Data Acquisition and Preprocessing

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 81003-D 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. *In vivo* 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% modulation transfer function (MTF), meaning that spatial frequencies of 1 line pair per mm are depicted with 50% contrast, 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 3-D Slicer tool [34]–[36], 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.

5680 slices of 13 teeth were selected for the training sets, and four other teeth (an incisor, a premolar, and two molars) provided 1824 slices for the test sets. In spite of the small number of teeth, the large variability of the slices allowed more precise measurements on a greater set of independent 2-D images. From hereon, the training set of LR CBCT images and HR μCT images will be denoted by TR_L and TR_H, and the corresponding test sets by TE_L and TE_H.

The CBCT and μCT images were uniformly normalized using the highest and lowest pixel values found in the TR_L and TR_H sets accordingly.

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. 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 images were renormalized between 0 and 1. It was qualitatively and quantitatively investigated how the SR algorithms handle this difference in noise patterns, and how do they perform after background removal.

B. Reconstruction-Based Deconvolution Methods

For evaluating the quality of our proposed SR method, our results are compared to a recent reconstruction-based SR

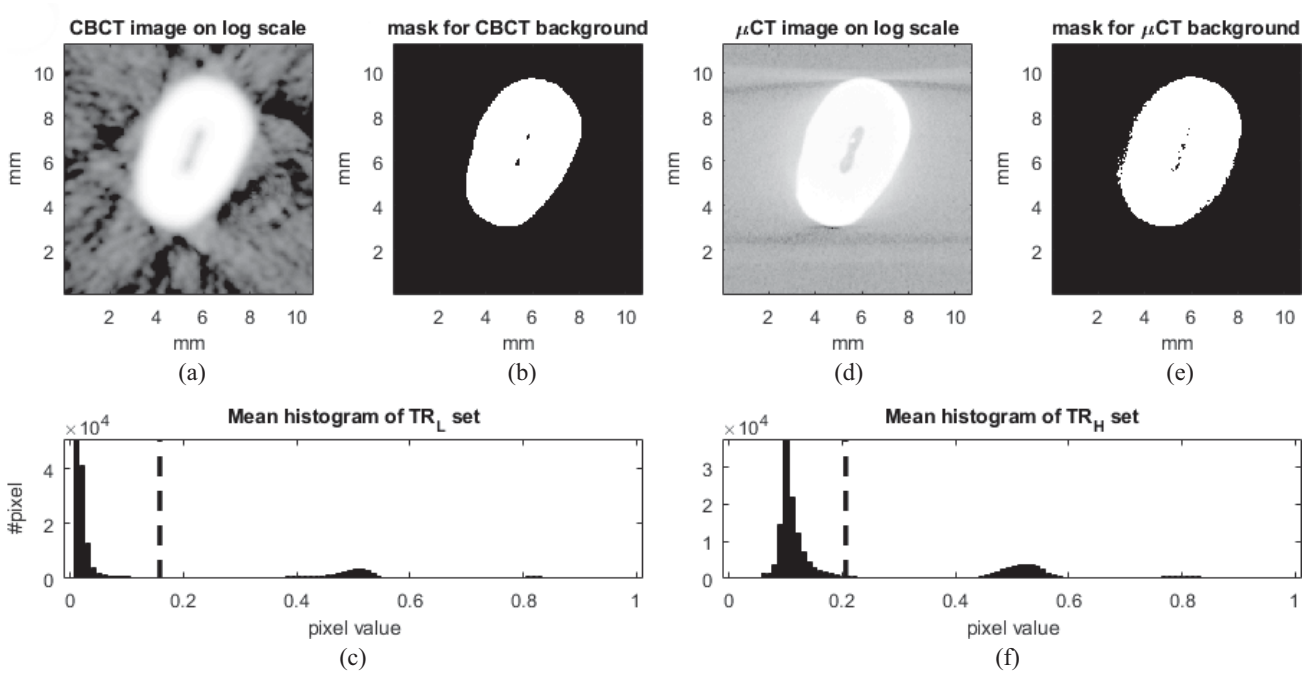

Fig. 1. Background artifacts. (a) Background artifacts on the CBCT image on log scale. (b) Mask of the tooth from (a). (c) Mean histogram of CBCT images from TR_L, with the global threshold used for background removal (dashed line). The same threshold was used for masking on (b). (d) Background artifacts on the μCT image on log scale. (e) Mask of the tooth from (d). (f) Mean histogram of μCT images from TR_H, with the global threshold used for background removal (dashed line). The same threshold was used for masking on (e).

method (SRR), implemented in MATLAB [37]. It should be noted that such approaches are among the most popular SR techniques. They assume that the observed LR image can be thought of as a noisy, blurred, and downsampled version of the HR image. The blurring effect is generally modeled as a convolution with a spatially invariant point-spread function (PSF). In practice the PSF is unknown, so it must be measured or estimated. Its estimation for experimental data is a very difficult task and is solved empirically in many existing works (see [38] for an example). In this paper, the CBCT images are assumed as low-pass filtered versions of the ideal μCT images. The PSFs were estimated using direct inverse filtering from each pair of the training CBCT- μCT images, where the constant λ was used to avoid dividing by 0. Finally, PSF-averaging over all the samples was processed to reduce noise. The PSF was thus obtained as

$$\mathrm{PSF} = \frac{1}{|\mathrm{TR}_L|} \sum_{k \in \mathrm{TR}_L} \frac{\mathcal{F}(\mathrm{TR}_L(k))}{\mathcal{F}(\mathrm{TR}_H(k)) + \lambda \cdot J}. \qquad (1)$$

The images are all resized to a common size. In (1) $\mathcal{F}$ denotes the Fourier transform operator, $|\mathrm{TR}_H|$ is the cardinality of the set, k is the training image index, 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. The ill-posedness of the inverse problem regarding the estimation of the HR image from its LR counterpart is overcome by incorporating regularization in the reconstruction process. In this paper two regularization terms are considered, namely the ℓ_2-norm and TV. The first was shown in [37] to lead to an inverse problem that can be solved analytically by exploiting particular properties of the downsampling and blurring operators. The second is well-known to promote piece-wise constant solutions, thus it was adapted to the application addressed in this paper.

C. Realizations of the CNN

The neural networks were realized using the open-access deep learning framework TensorFlow 1.3.0 [39], running with an NVIDIA GK210GL (Tesla K80 with 12 GB RAM) GPU. To investigate the potentials of these methods 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.

One of the investigated architectures was inspired by the U-net architecture [33] which is commonly used for domain-to-domain transformation, especially in medical imaging. It can also be modified to generate higher image dimensions than that of the original input image, but in this case the number of pixels was the same as in the CBCT and μCT images. We have implemented a structure with four successive downsampling layers on the original input image, continued by four upsampling steps which were implemented by transposed convolutions. At each size-level lateral connections concatenating

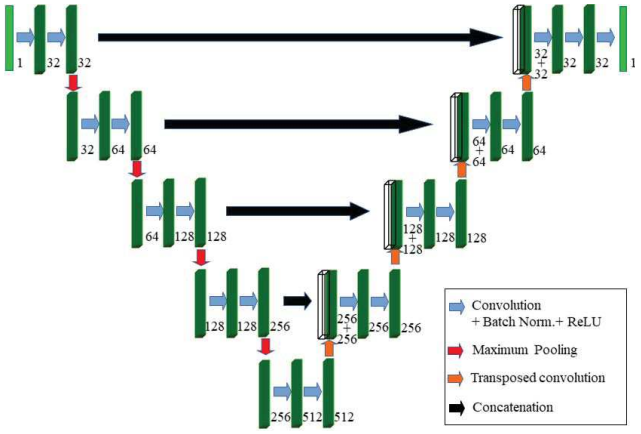

Fig. 2. Depiction of the U-net structure as a domain-to-domain transformation converting the input image with size 400x400x1 to an image of similar shape but different features. As it can be seen on the figure this structure is good to process local and global features together. The neurons in the deeper layers have larger and larger receptive fields. The numbers in the right bottom corner of the layers indicate the number of features stored.

the downsampled image features to the upsampled ones were also made. In each layer a combination of convolution, batch normalization [40] and rectified linear units (ReLU) was twice employed. Leaky implementations of ReLUs have been shown to provide higher accuracy and avoid the dying ReLU problem by providing a nonzero gradient for the constantly inactive neurons in the network [41]. The function of the leaky-ReLU (LReLU) is the following:

$$\mathrm{LReLU}(x) = \begin{cases} x & \text{if } x \geq 0 \\ \alpha x & \text{if } x < 0 \end{cases} \tag{2}$$

where x is the input response coming from the neuron and 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; in our case, it was set to 10^{-3}.

The number of convolutions—different features—were 32, 64, 128, 256 in the downsampling layers and 256, 128, 64, 32 in the upsampling layers. In the lowest resolution two convolutions with 512 features were also used.

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 [42]. Therefore, the size of all the kernels employed here was 3×3.

A detailed depiction of our architecture can be seen in Fig. 2.

Our second architecture for image enhancement was motivated by the subpixel networks implemented by Shi *et al.* [30], where deconvolution is realized as a tiling operator, instead of transposed convolutions [30]. We have implemented a commonly used six layer CNN structure containing an alteration of convolution, ReLU, and pooling operations in each layer, with 16, 32, 32, 64, 64, 4 convolutions, 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 size of the max-pooling kernels was

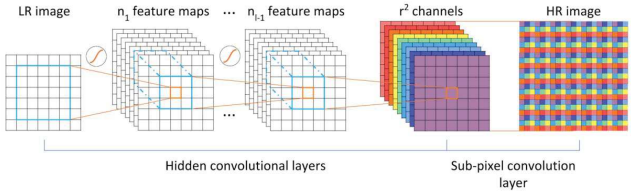

Fig. 3. Depiction of the retiling (depth-to-space) operation which we have also investigated for enhancing image quality. The image was taken from [30], showing an upsampling factor of three.

3×3 in all cases. The retiling operation (RT) that rearranges the elements of an $H \times W \times Dr^2$ tensor I to a tensor with a shape of $(rH \times rW \times D)$—and as such is responsible for the upscaling—can be defined as

$$\begin{aligned} \mathrm{RT}\{I\}(x, y, d) = I(\lfloor x/r \rfloor, \lfloor y/r \rfloor, D \cdot r \cdot \mathrm{mod}(y, r) \\ + D \cdot \mathrm{mod}(x, r) + d) \end{aligned} \tag{3}$$

where x, y, and d are the width, height, and depth indices of the input image, r is the upsampling factor (in our (in our case 2), D is the input depth of the image, $\lfloor \cdot \rfloor$ is the modulo operation. The depiction of this retiling can be seen in Fig. 3.

For training the networks on the 5680 slices of the TR_L and TR_H sets, the ADAM optimizer algorithm [43] 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 [44] and there were no significant differences in training depending on the weight and parameter settings. Similarly the initial timestep of the used ADAM optimizer did not have effect on overall reconstruction accuracy of the network. At each iteration of the training a random subset, batch of images was used to fit all computations in the memory of the GPU. We have used batches of 64 images for the CNN architecture and batches of 16 images for the U-net structure.

The concept of a loss function needs to be introduced, which is the error between two pixels backpropagated to improve the weights of the network with each iteration. The so-called ℓ_1 loss is the absolute difference, while the ℓ_2 loss is the squared difference (notice the analogy with the ℓ_1- and ℓ_2-norm). 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 [45]. On the other hand, ℓ_1 is only once differentiable, as opposed to ℓ_2. Here, a modified version of the Huber loss [46] was implemented, which combines the advantages of the two loss functions, helping the network to avoid local minima during training. The twice differentiable and smoother loss function ℓ_{1s} is

$$\ell_{1s}(O, G) = \begin{cases} |O - G| & \text{if } |O - G| \geq 1 \\ (O - G)^2 & \text{if } |O - G| < 1 \end{cases} \tag{4}$$

where O is the output image of the network and G is the desired output, the ground truth image. The loss function is then averaged over the entire aforementioned batch of images to yield an error that is then backpropagated. The networks may also be trained according to other metrics, as long as

they are fully differentiable. Note that for image normalization for the first layers and also between the layers we have used batch normalization [40]. This method ensures that input data in training batches is transformed to zero mean and unit variance. This means that those images and regions where a larger variance appeared fall into the $|O-G| \geq 1$ region. To the best of our knowledge this method is the most commonly used normalization method for deep learning image applications.

The structure of the two networks along with the algorithms and chosen parameters used for the training can be found on GitHub at: https://github.com/horan85/dentalsuperresolution.

D. Metrics

For the evaluation of similarity between corresponding 2-D images, we used the same metrics as in an earlier deep learning SR work [31]. Due to the complexity of some of the expressions, we limit ourselves to a brief description of the metrics, and refer the interested reader to the accompanying citations.

A simple and widely used measure is the mean squared error (MSE) calculated by averaging the squared differences of the reference and distorted image pixels. The peak signal-to-noise ratio (PSNR) is calculated by dividing the dynamic range by the MSE. These metrics, however, do not necessarily correspond to the perception of the human observer. The structure similarity index (SSI) was designed to better reflect subjective evaluation [47]. It combines the luminance, contrast, and structural measures, and can be calculated for single pixels considering small neighborhoods, or for the whole image as an average of the single values. The information fidelity criterion (IFC) quantifies the mutual information between two images, correlating with the perceptual quality [48]. The effect of frequency distortion and additive noise is estimated using the noise quality measure (NQM) [49]. These methods are all reference-based, meaning that a ground truth image—μCT in our case—is needed for the evaluation.

The enhanced images were also compared as 3-D volumes. The canal root was segmented from the 3-D volume using a dedicated adaptive local thresholding described in [50]. For visually showing the segmentation results, the software MeVisLab [51] was used. The segmentation results were analyzed quantitatively as well. For each root, the canal area and the Feret's diameter were estimated for all the radicular axial reconstructions, as suggested in [50]. The Feret's diameter defines the longest distance between two parallel straight lines that are tangent to the shape.

The comparison is first evaluated using the method of Bland and Altman through the bias (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 [52].

III. Results

A. Evolution of the Loss Function

Fig. 4 shows a comparison of convergences regarding the loss using the ℓ_1 (upper plot) and ℓ_{1s} (lower plot) functions.

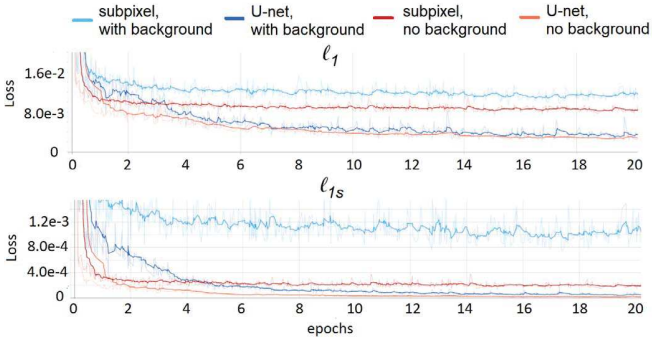

Fig. 4. Loss of the different networks during training according to the ℓ_1 and ℓ_{1s} metrics. To help visualize the general trends without the short-time randomness of the training algorithm, exponentially smoothed values are shown in dark, and the original values are plotted in semi-transparent colors.

The network was trained for 20 epochs, the loss converged and did not change significantly after ten 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 image quality metrics. When applying background removal, the loss values of the networks decrease for both loss functions (Fig. 4). Thus, background removal helped both networks to decrease the reconstruction error and to speed up the convergence.

B. Effect of Background Noise

The effect of the background noise was investigated qualitatively and quantitatively. The four SR algorithms—SRR-ℓ_2, SRR-TV, and CNNs with the subpixel and U-net architecture—were used on the test set, TE_L. An example slice can be seen in Fig. 5 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 remained 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.

The quantitative effect of the background noise can be seen in Table I. 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 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.

TABLE I
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	with background	CBCT	SRR:ℓ_2	SRR:TV	CNN:U-net	CCN: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**

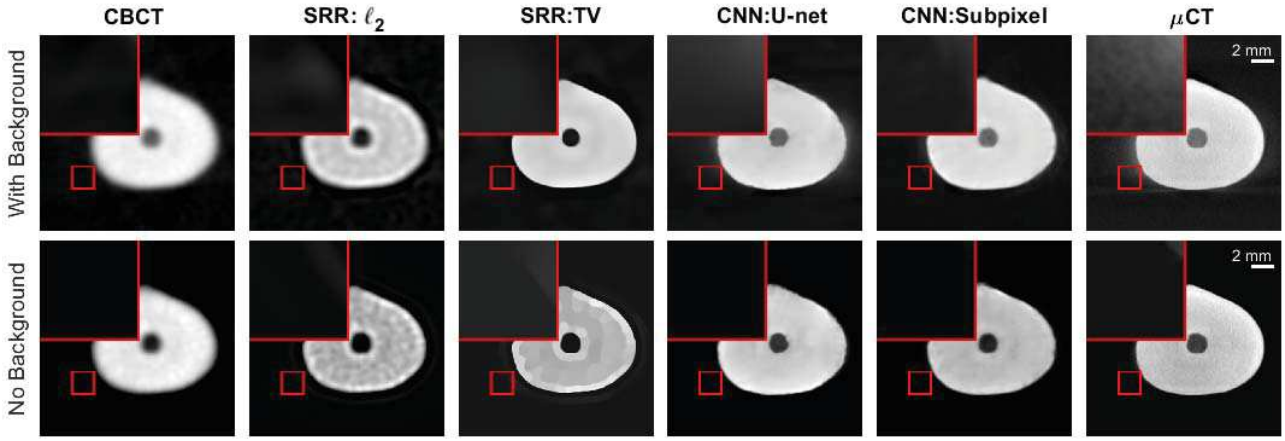

Fig. 5. Effect of background removal on noise amplification. The columns stand for the four different enhancement methods along with the original CBCT and μCT images. The upper row shows an example slice with intact background, while on the lower row background-removal was carried out. It can be seen 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].

When performing ℓ_2 and TV methods not all measures show improvement after background removal. Sometimes background removal seems to negatively affect reconstruction at the edges. It means that although we see 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.

C. Resolution Enhancement on Background-Removed Images

As we have qualitatively and quantitatively discussed the validity of background-removal, from hereon only the results obtained with the modified (without background) images will be examined. The values of Table I 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.

D. Comparison of 3-D Segmented Images

The quantitative results of the segmentation can be seen in Table II. 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 conspicuous 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 ℓ_2 Dice coefficient for the 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 3-D-visualization. The segmented canal structures of the CBCT-μCT and subpixel-μCT volume pairs were compared. Fig. 7 shows the two teeth

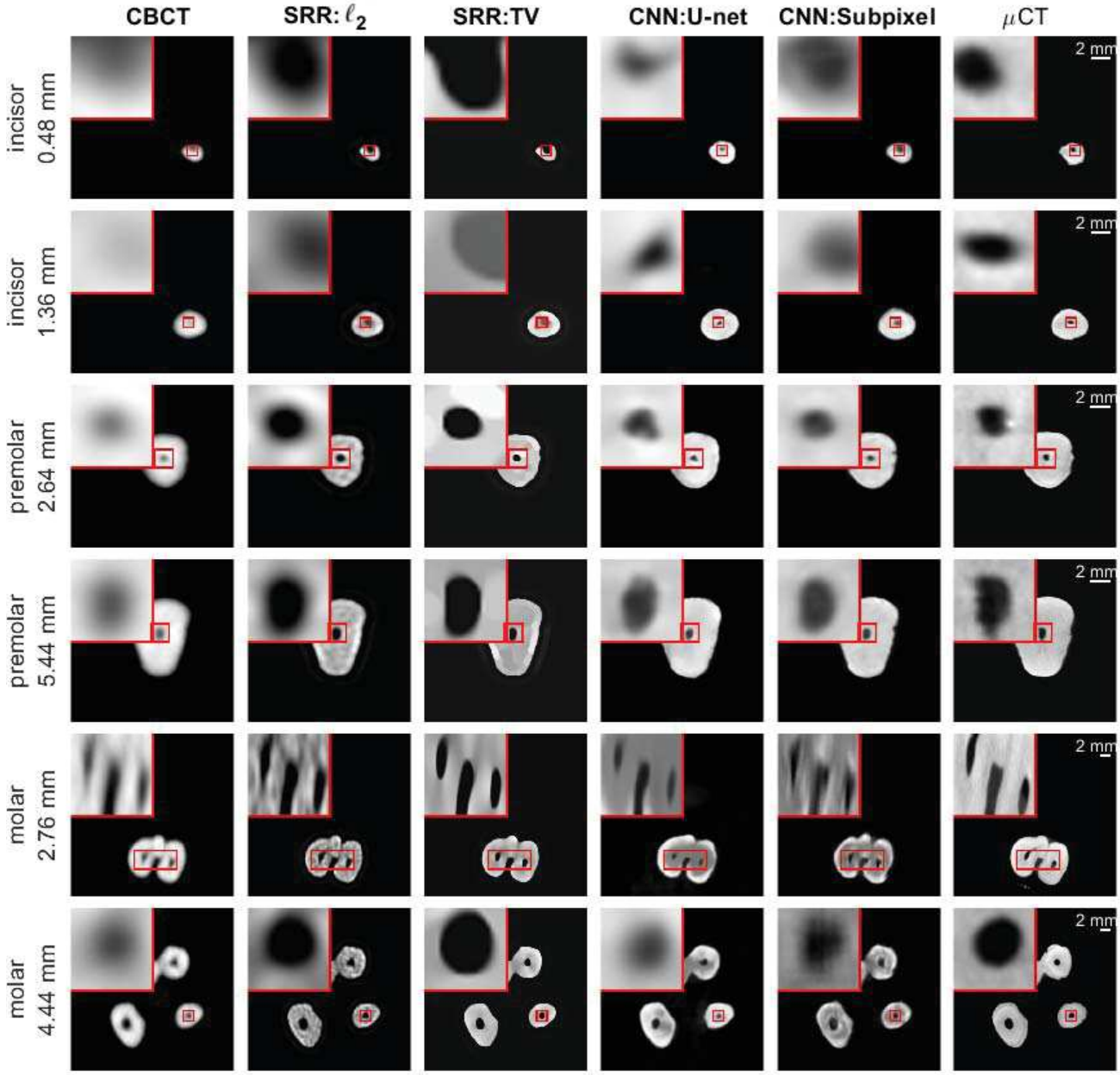

Fig. 6. Result of SR methods on different slices from the test set. On the left of the first column the type of the tooth and the depth of the slice from the apex of the root is displayed. The columns stand for the four enhancement methods along with the original CBCT and μCT images. The enhancement 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 I suggest. A 2 mm-scalebar is displayed on the μCT images. The display range is stretched to [0,1].

TABLE II
AVERAGE VALUES OF CANAL SEGMENTATION METRICS

Metric (compared to µCT volume)	CBCT	SRR:ℓ_2	SRR:TV	CNN:U-net	CCN:Subpixel
Mean of Differences - Area (mm^2)	0.0510	0.0674	0.0634	0.0500	**0.0327**
Mean of Differences - Feret (µm)	120.57	115.16	145.19	119.61	**114.26**
Difference of the Endodontic Volumes, x-µCT (%)	12.39%	12.25%	12.40%	10.12%	**6.07**%
Dice coefficient	0.8891	0.8852	0.8913	0.8998	**0.9101**

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.

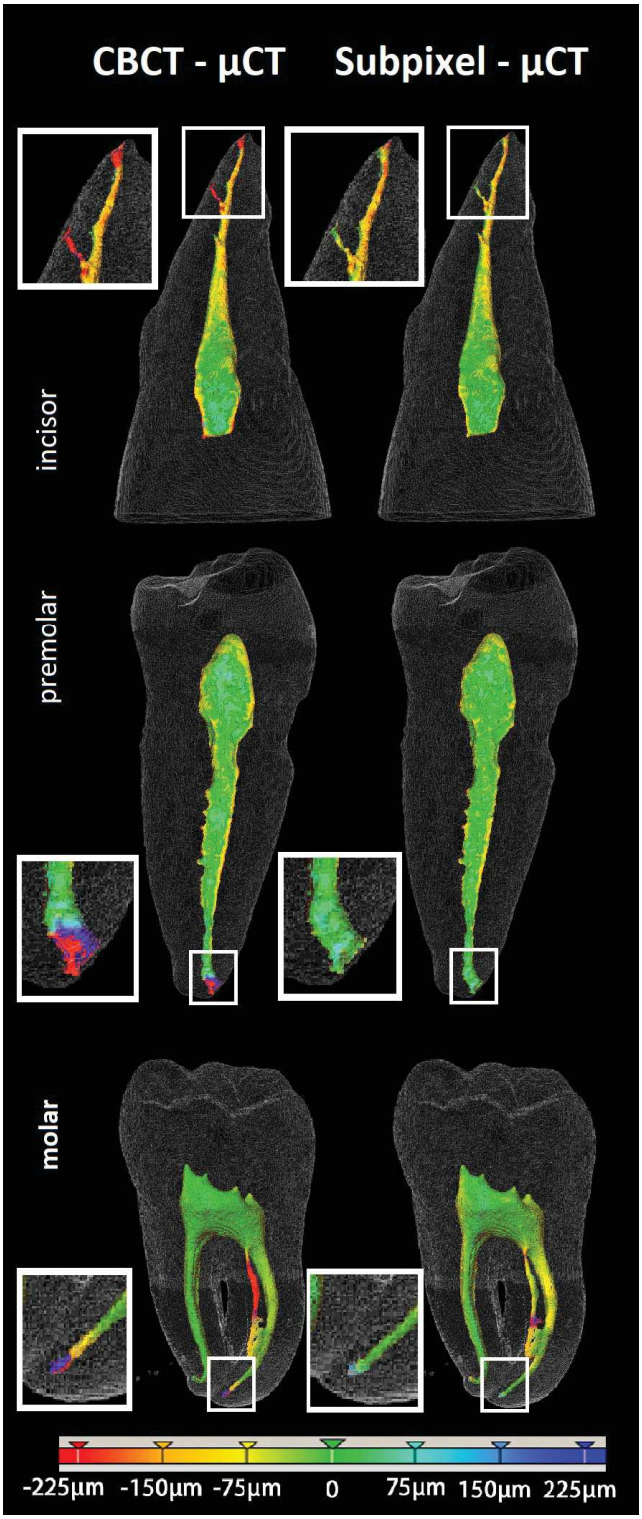

Fig. 7. Volumetric segmentation of the root canal on the test set (an upper incisor, a lower premolar tooth, and a lower molar). The colored area shows the difference between CBCT and μCT (on the left) and μCT segmentations. The highlighted areas show the apical end of the root, where the precision of the segmentation is more important during root canal treatment.

Similarly to Section III-B, 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.

IV. Conclusion

In this paper, two different deep-learning-based SR methods were implemented for dental CT 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 PNSR. In future work, the efficiency of different loss-functions and adversarial networks could be investigated in this regard. Further progress could be achieved by implementing networks with 3-D inputs, where information from neighboring slices could improve the training. Another interesting perspective of this paper is the application to phantom [53] or *in vivo* CBCT data, where the spatial resolution is further degraded compared to extracted teeth.

References

[1] O. A. Peters, "Current challenges and concepts in the preparation of root canal systems: A review," *J. Endodontics*, vol. 30, no. 8, pp. 559–567, 2004.

[2] 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," *Int. Endodontic J.*, vol. 40, no. 12, pp. 921–939, 2007.

[3] H. M. Eriksen, L.-L. Kirkevang, and K. Petersson, "Endodontic epidemiology and treatment outcome: General considerations," *Endodontic Topics*, vol. 2, no. 1, pp. 1–9, 2002.

[4] 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," *Int. J. Cardiol.*, vol. 148, no. 1, pp. 4–10, 2011.

[5] E. Cotti *et al.*, "Association of endodontic infection with detection of an initial lesion to the cardiovascular system," *J. Endodontics*, vol. 37, no. 12, pp. 1624–1629, 2011.

[6] J. J. Segura-Egea *et al.*, "Diabetes mellitus, periapical inflammation and endodontic treatment outcome," *Medicina Oral, Patología Oral Cirugía Bucal*, vol. 17, no. 2, p. e356, 2012.

[7] R. D. Astolphi *et al.*, "Periapical lesions decrease insulin signal and cause insulin resistance," *J. Endodontics*, vol. 39, no. 5, pp. 648–652, 2013.

[8] M. S. Gomes *et al.*, "Can apical periodontitis modify systemic levels of inflammatory markers? A systematic review and meta-analysis," *J. Endodontics*, vol. 39, no. 10, pp. 1205–1217, 2013.

[9] D. Brüllmann and R. K. W. Schulze, "Spatial resolution in CBCT machines for dental/maxillofacial applications-what do we know today?" *Dentomaxillofacial Radiol.*, vol. 44, no. 1, 2014, Art. no. 20140204.

[10] 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," *Anatomical Sci. Int.*, vol. 86, no. 3, pp. 146–150, Sep. 2011.

[11] K. Horner and S. Panel, "Cone beam CT for dental and maxillofacial radiology (evidence-based guidelines)," Eur. Commission Directorate Gen. Energy, Brussels, Belgium, Rep. 172, 2012. [Online]. Available: http://cordis.europa.eu/fp7/euratom/

[12] C.-Y. Yang, C. Ma, and M.-H. Yang, "Single-image super-resolution: A benchmark," in *Proc. 13th Eur. Conf. Comput. Vis. (ECCV)*, Sep. 2014, pp. 372–386.

[13] H. Chen *et al.*, "Low-dose CT via convolutional neural network," *Biomed. Opt. Exp.*, vol. 8, no. 2, pp. 679–694, Feb. 2017.

[14] C. Dong, C. C. Loy, K. He, and X. Tang, "Learning a deep convolutional network for image super-resolution," in *Proc. 13th Eur. Conf. Comput. Vis. (ECCV)*, Sep. 2014, pp. 184–199.

[15] G. Litjens *et al.*, "A survey on deep learning in medical image analysis," *Med. Image Anal.*, vol. 42, pp. 60–88, Dec. 2017.

[16] S. Sarraf and G. Tofighi, "Deep learning-based pipeline to recognize Alzheimer's disease using fMRI data," in *Proc. Future Technol. Conf. (FTC)*, 2016, pp. 816–820.

[17] H. I. Suk, S. W. Lee, and D. Shen, "Deep ensemble learning of sparse regression models for brain disease diagnosis," *Med. Image Anal.*, vol. 37, pp. 101–113, Apr. 2017.

[18] L. Zhao and K. Jia, "Multiscale CNNs for brain tumor segmentation and diagnosis," *Comput. Math. Methods Med.*, vol. 2016, Feb. 2016, Art. no. 8356294.

[19] 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.

[20] M. Ghafoorian *et al.*, "Deep multi-scale location-aware 3D convolutional neural networks for automated detection of lacunes of presumed vascular origin," *NeuroImage Clin.*, vol. 14, pp. 391–399, Jan. 2017.

[21] P. Yuehao *et al.*, "Brain tumor grading based on neural networks and convolutional neural networks," in *Proc. 37th Annu. Int. IEEE EMBC Conf.*, Milano, Italy, Aug. 2015, pp. 699–702.

[22] 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 *Proc. 1st Int. Workshop LABELS 2nd Int. Workshop DLMIA Held Conjunction MICCAI Deep Learn Data Labeling Med. Appl.*, Athens, Greece, Oct. 2016, pp. 95–110.

[23] 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 2nd Int. Workshop DLMIA Held Conjunction (MICCAI)*, Athens, Greece, Oct. 2016, pp. 170–178.

[24] 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.

[25] J. C. Ye, Y. Han, and E. Cha, "Deep convolutional framelets: A general deep learning framework for inverse problems," *SIAM J. Imag. Sci.*, vol. 11, no. 2, pp. 991–1048, 2018.

[26] F. E. A. Ali, Z. Nakao, Y.-W. Chen, K. Matsuo, and I. Ohkawa, "An adaptive backpropagation algorithm for limited-angle CT image reconstruction," *IEICE Trans. Fundam. Electron. Commun. Comput. Sci.*, vol. 83, no. 6, pp. 1049–1058, 2000.

[27] O. Oktay *et al.*, "Multi-input cardiac image super-resolution using convolutional neural networks," in *Proc. 19th MICCAI Int. Conf.*, Oct. 2016, pp. 246–254.

[28] S. Cengiz, M. D. C. Valdes-Hernandez, and E. Ozturk-Isik, "Super resolution convolutional neural networks for increasing spatial resolution of 1H magnetic resonance spectroscopic imaging," in *Proc. 21st MIUA Annu. Conf.*, Jul. 2017, pp. 641–650.

[29] Y. Zhang and M. An, "Deep learning- and transfer learning-based super resolution reconstruction from single medical image," *J. Healthcare Eng.*, vol. 2017, Jul. 2017, p. 20.

[30] W. Shi *et al.*, "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network," in *Proc. 29th IEEE Conf. CVPR*, Las Vegas, NV, USA, Jun. 2016, pp. 1874–1883.

[31] 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.

[32] J. Kim, J. K. Lee, and K. M. Lee, "Accurate image super-resolution using very deep convolutional networks," in *Proc. 29th IEEE Conf. CVPR*, Las Vegas, NV, USA, Jun. 2016, pp. 1646–1654.

[33] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proc. 18th Int. Conf. MICCAI*, Oct. 2015, pp. 234–241.

[34] 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*. New York, NY, USA: Springer, 2014, ch. 19, pp. 277–289.

[35] A. Fedorov *et al.*, "3D slicer as an image computing platform for the quantitative imaging network," *Magn. Resonance Imag.*, vol. 30, no. 9, pp. 1323–1341, Nov. 2012.

[36] (2017). *3D Slicer*. [Online]. Available: https://www.slicer.org/

[37] N. Zhao *et al.*, "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.

[38] A. Toma, L. Denis, B. Sixou, J.-B. Pialat, and F. Peyrin, "Total variation super-resolution for 3d trabecular bone micro-structure segmentation," in *Proc. IEEE 22nd Eur. Signal Process. Conf. (EUSIPCO)*, 2014, pp. 2220–2224.

[39] M. Abadi *et al.* (2015). *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. [Online]. Available: https://www.tensorflow.org/

[40] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. 32nd Int. Conf. Mach. Learn. (ICML)*, Lille, France, 2015, pp. 448–456.

[41] B. Xu, N. Wang, T. Chen, and M. Li, "Empirical evaluation of rectified activations in convolutional network," in *Proc. 32nd Int. Conf. Mach. Learn. Deep Learn. Workshop (ICML)*, 2015.

[42] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Las Vegas, NV, USA, 2016, pp. 2818–2826.

[43] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. 3rd Int. Conf. Learn. Represent. (ICLR)*, 2015.

[44] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proc. 13th Int. Conf. Artif. Intell. Stat.*, 2010, pp. 249–256.

[45] H. Zhao, O. Gallo, I. Frosio, and J. Kautz, "Loss functions for image restoration with neural networks," *IEEE Trans. Comput. Imag.*, vol. 3, no. 1, pp. 47–57, 2017.

[46] P. J. Huber, "The place of the L1-norm in robust estimation," *Comput. Stat. Data Anal.*, vol. 5, no. 4, pp. 255–262, Sep. 1987.

[47] 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.

[48] 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.

[49] 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.

[50] J. Michetti, A. Basarab, F. Diemer, and D. Kouame, "Comparison of an adaptive local thresholding method on CBCT and μCT endodontic images," *Phys. Med. Biol.*, vol. 63, no. 1, 2017, Art. no. 015020.

[51] (2017). *MeVisLab*. [Online]. Available: https://www.mevislab.de

[52] L. R. Dice, "Measures of the amount of ecologic association between species," *Ecology*, vol. 26, no. 3, pp. 297–302, 1945.

[53] 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 37th Annu. Int. Conf. Eng. Med. Biol. Soc. (EMBC)*, 2015, pp. 7905–7908.