Real Super-Resolution for Proximal Femur: Enhanced Computation of Structural Bone Metrics from Clinical CTs

Niklas C. Koser[1]*, Marten J. Finck[2]*, Felix N. von Brackel[3], Benjamin Ondruschka[4], Sören Pirk[2], Claus-C. Glüer[1], and

[1] i2Lab@Section Biomedical Imaging, Kiel University, University Hospital Schleswig-Holstein, Kiel {niklas.koser,glueer}@rad.uni-kiel.de
[2] Visual Computing and Artificial Intelligence, Kiel University, Kiel {mafi,sp}@informatik.uni-kiel.de
[3] Department of Osteology and Biomechanics, University Medical Center Hamburg–Eppendorf, Hamburg
[4] Institute of Legal Medicine, University Medical Center Hamburg-Eppendorf, Hamburg

Abstract. Fracture risk due to osteoporosis is a highly prevalent disease with costs in the European Union alone of 56 billion p.a.. Accurate assessment of the microarchitecture of the proximal femur (e.g., trabecular thickness, trabecular spacing, bone volume fraction) is essential for assessing bone strength and predicting fracture risk. High resolution (HR) CT provides the necessary spatial resolution. However, for best hip fracture risk assessment HR-CT imaging should be performed at the proximal femur but this would require an unacceptably high level of radiation dose. Therefore, we aimed to investigate whether deep learning based super-resolution (SR) models applied to low-resolution (LR) clinical CT images permit improved assessment of structural parameters.
In this study we adapted and optimized state-of-the-art model architectures to compare them in the context of CT-SR of the proximal femur. The dataset used consisted of pairs of clinical LR-CTs and HR-CTs of 50 individuals. This represents clinical reality and avoids bias of downsampling HR images to mimic LR images. Using automated preprocessing data is prepared for model training. We used three-stage template matching of point clouds to automatically extract the relevant regions of interest, from which metrics for bone microarchitecture were determined. We compared SRGAN, Real-ESRGAN+, LDM, and ResShift regarding improvement in structural assessment. We also tested whether 2.5D approaches –using multiple slices of the CT– are superior to 2D approaches. In terms of perceptual reconstruction, the ResShift 2.5D model outperforms the other SR models and achieves comparable results to the Real-ESRGAN+ architectures in the derivation of biomechanical properties.

Keywords: CT Super-Resolution · Bone Microstructure · Clinical CT · Osteoporosis · Deep Learning.

* Contributed equally to this work.
Code and data are available at: https://github.com/nkoser/Real-Super-Resolution

1 Introduction

Osteoporosis is a systemic skeletal disease characterized by reduced bone mass and microarchitectural structure of the bone [1, 21]. The trabecular bone structure is degraded faster than the compact cortical bone, which is why bones with a high proportion of spongiosa, such as the proximal femur, have an increased risk of osteoporotic fractures [1]. The prevalence of osteoporosis increases significantly with age. In view of demographic trends and increasing life expectancy, it is predicted that the incidence of osteoporotic fractures will continue to rise substantially in the future [1, 18, 21]. This development significantly impairs the quality of life of those affected and also represents a considerable financial burden for healthcare systems – in 2019 alone, 4.3 million fractures in the European Union resulted in costs exceeding 56 billion euros [1, 15, 21]. To prevent these consequences, an efficient prognosis and early diagnosis are crucial, as timely preventive measures in line with guidelines yield high success rates. According to the current recommendations of the World Health Organization (WHO), the diagnosis of osteoporosis is primarily based on determining bone mineral density (BMD) using Dual-Energy X-Ray Absorptiometry (DXA) [1, 21]. However, this standard metric only has a classification accuracy of around 65 % for assessing risk of hip fracture [11]. The scientific literature therefore recommends the integration of additional metrics that capture the microstructural bone properties. Parameters such as bone volume fraction (BV/TV), trabecular thickness (Tb.Th), trabecular spacing (Tb.Sp) and trabecular number (Tb.N) have been shown to be predictive of the mechanical strength of the bone and thus offer high diagnostic and prognostic potential for estimating fracture risk [1, 18–20].

High-resolution peripheral quantitative computed tomography (HRpQCT), here denoted as HR-CT, facilitates the acquisition of detailed images essential for analyzing the bone microstructure. However, it is currently mainly used in scientific settings because of high costs, a time-consuming procedure, and a comparatively small field of view. It is primarily limited to peripheral skeletal regions and would pose a high radiation exposure for *in vivo* imaging [1, 11, 19, 20]. To overcome these limitations, super-resolution (SR) methods could be applied to clinical CT images that are characterized by a lower spatial resolution (denoted as LR-CT). SR methods are used to increase the image resolution in order to achieve an approximation of the quality of HR images. Deep learning (DL) based approaches, which are capable of significantly improving image quality by training on extensive paired data sets of LR and HR images, are particularly promising. Recent studies show that these methods outperform common interpolation methods in terms of image sharpness, texture detail and structural accuracy [17]. A noteworthy class of algorithms are Generative Adversarial Networks (GANs), which generate realistic HR-images through competition between a generator and a discriminator network [16, 17, 23, 24]. More recently, diffusion models have emerged as a promising class of algorithms due to their stable training dynamics and ability to provide highly detailed reconstructions [22].

In the field of DL-SR imaging of the proximal femur, only two approaches have been introduced so far – and both of these studies exhibit methodological and data-related limitations that hinder their applicability in clinical practice. Chan and Rajapakse [4] trained an SR3 model based on 26 HR-CT volumes with an isotropic voxel spacing of 240 µm. The corresponding LR-CT volumes were generated by bicubic downsampling to 720 µm. Despite improvements over methods like SRGAN, the significance of these results is limited by the small dataset size and the use of synthetically generated LR volumes. These downsampled images may still retain aspects of the HR data that are absent in truly independent LR acquisitions, making them less representative of real clinical imaging conditions. Frazer et al. [10] developed a SRGAN using a paired data set of 10 HR-CT volumes (60 µm isotropic) and 10 LR-CT volumes (0.3125 x 0.3125 x 0.25 mm) acquired, however, with a special clinical CT scanner that is not representative of commonly available CT scanners, as acknowledged by the authors. Therefore, the significance of these results is also limited. Besides the small and unbalanced sample size, the scaling applied during registration possibly leading to distortions in the metrics is a limitation. Although both studies are promising, they do not use current state-of-the-art (SOTA) SR models. Furthermore, they are based on inadequate datasets that are limited in both size and clinical relevance, as they still contain too much detail in the LR-CTs. There is also a lack of standardized evaluation methods, which makes it difficult to directly compare the results.

This paper aims to address these limitations by proposing the following contributions: (1) We introduce the largest paired dataset to date, consisting of 50 HR-CTs and corresponding LR-CTs, with the LR-CTs recorded under real clinical conditions. This dataset enables a realistic evaluation of SR methods in a medical context and should be made available to the research community; (2) We adapt and test various current SOTA SR models on 2D and 2.5D slices of the described dataset, also including the recently released ResShift [27] model that has not been evaluated on medical image data before; (3) We propose an automated evaluation pipeline based on template matching of surface point clouds. This method enables a standardized selection of regions of interest (ROIs), ensuring a reproducible evaluation of SR models based on established metrics.

2 Methodology

2.1 Individuals, Dataset and Data Preprocessing

Our dataset consists of paired CT images of 50 patients, including 31 male and 19 female individuals. The age of the patients ranges from 22 to 89 years, with an average age of 60.8 ± 17.5 years. Body height varies from 1.43 m to 2.00 m, with a mean of 1.73 m ± 0.11 m, while body weight ranges from 44 kg to 123 kg, with an average of 81.6 kg ± 19.4 kg. The LR-CTs were acquired post mortem *in situ* as whole-body scans with a Philips Incisive CT. Image acquisition was performed with anisotropic voxel spacing of 0.98 mm × 0.98 mm × 0.65 mm, peak kilo voltage output (kVp) of 120 kV, tube current of 113 - 130 mA, and an UB filter. In the femur region, these CT scans exhibit a mean Signal-to-Noise Ratio (SNR) of 0.93 and a mean contrast ratio of 2.43. Subsequently, the HR-CTs were acquired using an HR-pQCT (Xtreme CT, SCANCO Medical AG, Brüttisellen, Switzerland). The left proximal femur was scanned *ex situ* with an isotropic voxel resolution of 82 µm, a kVp of 59.4 kV, 900 µA, and reconstructed with the Shepp-Logan kernel. This study was approved by the ethics committee of the Hamburg chamber of physicians (Reference Number: WF-057/21) and conducted in accordance with the ethical standards of the Declaration of Helsinki of 1964 and its later amendments.

In order to train neural networks for SR, semi-automatic preprocessing of our dataset is required to generate corresponding volume pairs. The following processing steps are performed for each pair of LR-CT and HR-CT. First, both DICOM volumes are loaded with the open source software 3D Slicer [8]. For initial adjustment, the LR-CT is roughly aligned with the HR-CT by manual translation and rotation. This is followed by precise registration using the BRAINS General Registration Module [14] with a rigid six degrees of freedom transformation model. The LR-CT serves as a moving volume and the HR-CT as a fixed volume. After registration, the LR-CT is resampled onto the isotropic voxel grid of the HR-CT using a specially developed 3D slicer module, which performs resampling with SimpleITK and a B-spline interpolation, and is subsequently cropped to match the spatial dimensions of the HR-CT. As only the LR-CT is transformed, the structural integrity of the HR-CT is preserved, ensuring that no relevant information is lost. This procedure is independent of the voxel spacing of the LR clinical CT scanner. Therefore, it is also transferable to other scanners. A segmentation mask is required for the further preprocessing steps of the volumes exported as DICOM. To automatically compute these masks, we use a 3D nnUNet [13] trained on a subset of the Total Segmentator dataset containing a total of 1,204 labelled CT images [26]. For this purpose, 102 CTs and their corresponding segmentation masks were first cropped to the relevant areas of the left and right femora. The resulting 204 volumes with an isotropic voxel spacing of 1 mm were then used as the basis for training. The trained nnUNet enables generating masks from downsampled LR-CTs of our dataset and upsamples them to the original resolution with voxel spacing of 82 µm. The value ranges were normalized to the range $[0, 1]$ and clipped. Finally, the mask is applied so that only the femur bone and no other structures (e.g., tissue or neighboring bones) is present in the volume.

2.2 Model Architectures

To enhance the resolution and reconstruction of anatomical structures in LR-CTs of the proximal femur, we adapted and compared four deep learning SR models (SRGAN [16], Real-ESRGAN+ [23], Latent Diffusion Model (LDM) [22] and ResShift [27]) both in 2D and 2.5D. The latter also includes four neighboring slices assuming that the contained structural information represents an advantage for the reconstruction of the considered slice.

SRGAN. Since Frazer et. al. [10] were able to achieve promising results with an SRGAN, we aim to evaluate the capabilities of this model architecture trained on real LR-CTs of the proximal femur. Due to the voxel-to-voxel correspondence between the LR and HR-CTs, we removed the upsampling layer. Furthermore, we added four additional residual blocks into the generator, given that the dimensions of the input pixels exceed those of the original SRGAN. No modifications to the model were necessary for 2.5D, because the neighboring slices were integrated as additional channels.

Real-ESRGAN+. A strong limitation of SRGAN is its inability to reproduce finer structures in LR images [16]. To address these limitations Wang et al. developed ESRGAN [24] and Real-ESRGAN+ [23]. The latter approach not only focus on the reconstruction of fine details, but also introduces a process to eliminate undesirable artifacts that arise during the reconstruction process. Due to the larger input size compared to Real-ESRGAN+, we added three additional Residual-in-Residual Dense Blocks. Because of the low image sharpness of the LR-CTs and the different acquisition conditions compared to the HR-CT in our dataset, the default degradation process is omitted. No further adjustments were made to the original model regarding pixel unshuffling and the use of a unsharp masking filter to mitigate overshooting artifacts. The 2.5D input was processed in the same way as for the 2D input.

LDM. Due to the known challenges of training GANs [22], we use a LDM as another class of SR model algorithms. We deliberately decided against the SR3 model used by Chan et al. [4], as LDMs offers both higher efficiency and the ability to generate more realistic textures [22]. A LDM uses a variational autoencoder (VAE) to transform the HR image into a latent space with dimension that corresponds to those of the LR image. In our specific case, we apply pixel unshuffling to transform the LR-CT image to the dimension of the HR latent space $(6, 64, 64)$ without information loss. In the 2.5D case, the number of channels was increased to accommodate the requirements of pixel unshuffling.

ResShift. In contrast to LDMs, ResShift [27] is capable of generating visually plausible images within 15 iterations, facilitated by the formation of a Markov chain between HR and LR images. This is achieved by shifting the residual – the difference between the HR and LR images – which enhances the efficiency of transitions. ResShift demonstrates comparable or even superior performance compared to other SOTA SR methods across various applications [27]. Building on these findings, our study is the first to apply and evaluate ResShift on medical data, employing it as the fourth SR model in our experimental setup. To enable even faster training of ResShift in the latent space, we utilize a variational autoencoder (VAE) trained on HR-CTs. The VAE is also applied to LR-CTs to ensure alignment within the shared latent space. Additionally, a feature extractor is employed on the original LR-CTs to transmit LR features as conditioning inputs to the Swin-Unet [3]. In the 2.5D approach, the features of adjacent slices – extracted using the feature extractor adapted for five channels – are likewise supplied as conditioning inputs to the Swin-Unet.

3 Experiments and Results

3.1 Experimental Setup

Evaluation. Since the structural relationships are of primary importance in the reconstruction of the slices, the calculation of Structural Similarity Index (SSIM) and Peak Signal-to-Noise Ratio (PSNR) is performed on the entire slices and are then averaged over the volume [25]. As SSIM and PSNR are sometimes unreliable for blurred images, we also record the Learned Perceptual Image Patch Similarity (LPIPS) [28] and the Gradient-SSIM (GSSIM) [5]. The average inference time (T/V) is also measured. To consistently and automatically extract structural bone metrics from the same anatomical region, independent of patient-specific variations and without manual ROI selection, we developed a multi-stage template matching approach based on surface point clouds. Bone volume, defined as the sum of bone voxels within the binary mask, was used to select three templates, which were uniformly sampled to ensure anatomical diversity. The three selected femur templates were segmented into anatomical regions (head, neck, trochanter, shaft) using 3D Slicer. The segmentation masks were converted into triangle meshes using Open3D [29], from which 10,000 labeled points were uniformly sampled as point clouds. The following procedure was then applied to all bone volume pairs (LR-CT/HR-CT or SR-CT/HR-CT): (1) The appropriate template is selected for both volumes and their segmentation mask. For each CT volume, a binary bone mask is then generated from a nnUNet trained on HR-CTs; (2) A surface point cloud is generated from the segmentation mask in the same way as the templates. We then apply a three-stage point cloud registration process. First, the point cloud centers are aligned by translation. This is followed in the second step by a rough registration using RANSAC [9]. Finally, a more precise fine adjustment is applied using ICP with scaling [2]. Then the labels of the template point cloud are transferred to the transformed point cloud so that a complete segmentation into the four anatomical regions is achieved; (3) The center of all associated points is calculated for the anatomical regions of the head, neck, and trochanter. 3D-ROIs with an edge length of 128 voxels are extracted around these reference points from the bone structure masks; (4) To evaluate the bone structure, BV/TV, Tb.Th, Tb.Sp, and Tb.N are calculated from the extracted ROIs using the Hildebrand algorithm [12].

Implementation Details. The dataset is divided into training, validation, and testing, corresponding to a 60/20/20 split. Due to the size of the volume data, 2D and 2.5D patches with a size of 512×512 are extracted from the axial slices. Sampling is centered around the voxel located inside the femur. Rotation and flipping are used to augment the data and ensure robustness.

All models are trained with Adam optimizer and a batch size of 64, with each model processing 10,000 samples per epoch. The NVIDIA H100 NVL GPU with 96 GB VRAM was used for this purpose. To train the SRGAN, the generator is first pre-trained with a learning rate of 1e-4 (multi-step learning rate scheduler) for 45 epochs in order to avoid overfitting the discriminator during

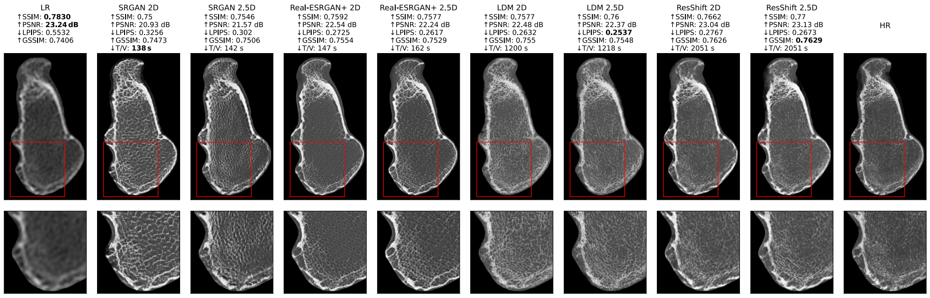

Fig. 1. Comparison of the tested SR models based on their reconstruction performance using an axial slice and an enlarged view (30 mm edge length).

the subsequent GAN training. The actual GAN training then takes place over 75 epochs, whereby the discriminator is optimized with a learning rate of 1e-6. The same training procedure is used for the Real-ESRGAN+. The generator is first pre-trained with a learning rate of 2e-4 over 20 epochs (constant learning rate scheduler), followed by adversarial GAN training over 60 epochs, whereby the discriminator is optimized with a learning rate of 1e-4. The LDM model was trained over 100 epochs with a learning rate of 1e-5 where forward diffusion was carried out with the DDPM method ($T = 1000$) and inference with the DDIM method ($T = 30$). The VAE was pre-trained over 75 epochs using a perceptual loss and a mean squared error (MSE) loss. A patch discriminator was then used for adversarial training (75 epochs, learning rate of 45e-6). The ResShift model was trained over 150 epochs with the same learning rate, using 8 diffusion steps and a Cosine scheduler. Our code will be shared publicly upon acceptance.

3.2 Comparative Analysis

The results used to assess the reconstruction performance of the SR models are presented in Figure 1, exemplified by an axial slice from a test volume. Additionally, the average SSIM, PSNR, LPIPS and GSSIM values across all slices and volumes, along with the mean T/V, are reported. Qualitative analyses show that models based on GAN architectures generate symmetric recurrent structures as trabecular meshwork. SRGAN 2D in particular hallucinates microstructures that are not present in HR-CT. The LDM, on the other hand, is not able to correctly reproduce prominent structures such as the cortex and also shows only minor local differences in the trabecular meshwork. The most convincing model is the ResShift model, especially when trained on 2.5D data, as it can reconstruct both the cortex and the trabecular distribution well. For the other three models, the added value of 2.5D is marginal. The analysis of the quantitative metrics shows that SSIM and PSNR have the highest values for the LR-CT and thus do not match the qualitative results. The 2.5D LDM model achieves the best LPIPS value. The highest agreement with the visually assessed image quality is shown

Table 1. Mean, standard deviation and Pearson correlation coefficient (r) of the bone microstructure metrics. Tb.Th and Tb.Sp are in mm, Tb.N in mm^{-1}.

		SRGAN		Real-ESRGAN+		LDM		ResShift		
	LR	2D	2.5D	2D	2.5D	2D	2.5D	2D	2.5D	HR
BV/TV$_{\mathrm{h}}$	.15±.2 r=.5	.54±.14 r=.35	.46±.15 r=.42	.43±.16 r=**.73**	.41±.14 r=.71	.62±.07 r=.29	.64±.08 r=.37	0.59±0.1 r=.64	.62±.1 r=**.73**	.56±.06
Tb.Th$_{\mathrm{h}}$	.86±.53 r=.33	.6±.08 r=.64	.63±.1 r=.81	.51±.12 r=.77	.5±.1 r=.8	.54±.06 r=.82	.57±.07 r=.74	.63±.11 r=**.86**	.63±.11 r=**.86**	.58±.08
Tb.Sp$_{\mathrm{h}}$	5.1±3.8 r=0	.63±.15 r=.13	.89±.49 r=−.08	.65±.14 r=.2	.65±.12 r=**.33**	.41±.06 r=−.22	.41±.06 r=−.1	.5±.09 r=−.06	.47±.07 r=.07	.64±.05
Tb.N$_{\mathrm{h}}$	.21±.09 r=−.58	.82±.09 r=.22	.69±.15 r=−.15	.86±.06 r=.21	.88±.05 r=.47	1.1±.07 r=.74	1±.06 r=**.79**	.89±.09 r=.67	.92±.09 r=.62	.82±.06
BV/TV$_{\mathrm{n}}$	.01±.01 r=.35	.31±.2 r=−.03	.25±.2 r=.21	.05±.04 r=.82	.06±.04 r=.77	.1±.07 r=.64	.12±.07 r=.83	.06±.05 r=.81	.06±.04 r=**.85**	.11±.1
Tb.Th$_{\mathrm{n}}$	.79±1.18 r=.95	.46±.11 r=.85	.46±.14 r=.92	.4±.32 r=**.97**	.36±.2 r=**.97**	.32±.21 r=**.97**	.34±.19 r=**.97**	.36±.27 r=**.97**	.35±.29 r=**.97**	.35±.14
Tb.Sp$_{\mathrm{n}}$	8.4±8.3 r=−.41	.99±.53 r=−.28	1.4±.95 r=.39	2.8±1.9 r=.72	2.1±1.3 r=**.93**	1.3±.78 r=−.2	1±.19 r=.51	2.3±.83 r=.49	2.2±1 r=.74	3.5±6
Tb.N$_{\mathrm{n}}$	.07±.07 r=.37	.76±.24 r=−.27	.66±.29 r=.13	.41±.18 r=.68	.49±.17 r=**.75**	.72±.24 r=−.18	.75±.13 r=.51	.42±.15 r=.62	.44±.16 r=.74	.53±.25
BV/TV$_{\mathrm{t}}$	0±.01 r=.27	.27±.2 r=−.24	.22±.19 r=.13	.03±.04 r=**.76**	.04±.03 r=.63	.05±.03 r=.08	.05±.04 r=.19	.02±.01 r=.06	.02±.02 r=.71	.11±.09
Tb.Th$_{\mathrm{t}}$	.22±.24 r=.26	.4±.07 r=−.18	.37±.06 r=.34	.26±.04 r=**.83**	.27±.02 r=.75	.22±.02 r=.15	.23±.02 r=.58	.22±.03 r=.79	.21±.02 r=.79	.3±.05
Tb.Sp$_{\mathrm{t}}$	6.9±8.3 r=−.19	1±.64 r=−.15	1.5±1.2 r=−.17	3.1±2.2 r=**.83**	1.9±.76 r=.6	1.3±.5 r=−.08	1.4±.41 r=−.21	2.8±1 r=.58	2.9±1.5 r=.48	1.6±.8
Tb.N$_{\mathrm{t}}$	.05±.07 r=.26	.78±.23 r=−.18	.7±.3 r=−.04	.4±.2 r=**.66**	.51±.16 r=.6	.72±.19 r=.01	.66±.16 r=−.09	.37±.12 r=.58	.4±.19 r=.62	.6±.18

by the GSSIM, which is highest for ResShift 2.5D at 0.7629 and lowest for the LR-CT at 0.7406. The GAN-based models have a significantly shorter inference time compared to the diffusion models. While the SRGAN 2D generates an SR volume in 138 seconds on average, the LDM 2D requires 1,200 seconds. The metrics of bone microstructure are compared for all models in Table 1.

The analysis of the correlation coefficients shows that the ResShift 2.5D and the Real-ESRGAN+ architectures in particular are superior to the other SR models examined in the reconstruction of bone microstructures. Both models, when applied in a 2.5D configuration, demonstrated statistically significant improvements over metrics derived from the LR-CT scans in 8 out of 12 evaluation metrics, as determined by Friedman's test followed by a Conover post-hoc analysis [6, 7]. In contrast, the SRGAN achieved the lowest correlation with the HR-CTs. In most cases, all models provide a more accurate estimation of structural bone metrics.

4 Conclusion, Limitations and Future Work

In this paper we adapted and compared different SOTA SR model architectures – trained on a newly introduced dataset reflecting the clinical reality – for improving the calculation of bone microstructure metrics. Our analysis demonstrates, that ResShift 2.5D outperforms existing model architectures regarding perceptual reconstruction and shows similar results to the Real-ESRGAN+ when deriving structural bone metrics. Additionally, our introduced evaluation pipeline enables standardized comparisons for other approaches. Although our results are promising, we also identified the following limitations: (1) The LR-CTs have limited resolution due to the real clinical setting leading to a loss of information and possible hallucination of bone microstructure; (2) Our method has only been validated with one clinical CT scanner; (3) Since the training is performed exclusively with axial slices, the reconstruction quality in the other two dimensions is less precise; (4) The quality of the results in both preprocessing and evaluation depends significantly on the accuracy of the binary masks. Given the current state, there are several avenues for future work, including the assessment with other metrics, the training on 3D patches and the validation through a suitable downstream task.

Acknowledgments. This work has been financially supported by the Federal Ministry of Education and Research, Germany (grant 01EC1908A). Furthermore, this work was supported by the IDIR-Project (Digital Implant Research), a cooperation financed by Kiel University, University Hospital Schleswig-Holstein and Helmholtz Zentrum Hereon.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Bartl, R. (ed.): Osteoporose: Prävention · Diagnostik · Therapie. Georg Thieme Verlag, Stuttgart, 01 edn. (2010)
2. Besl, P.J., McKay, N.D.: Method for registration of 3-d shapes. In: Sensor fusion IV: control paradigms and data structures. vol. 1611, pp. 586–606. Spie (1992)
3. Cao, H., Wang, Y., Chen, J., Jiang, D., Zhang, X., Tian, Q., Wang, M.: Swin-unet: Unet-like pure transformer for medical image segmentation. In: Computer Vision – ECCV 2022 Workshops. pp. 205–218. Springer Nature Switzerland (2023)
4. Chan, T.J., Rajapakse, C.S.: A super-resolution diffusion model for recovering bone microstructure from ct images. Radiology: Artificial Intelligence **5**(6), e220251 (2023)
5. Chen, G.h., Yang, C.l., Xie, S.l.: Gradient-based structural similarity for image quality assessment. In: 2006 International Conference on Image Processing. pp. 2929–2932 (2006)
6. Conover, W.J.: Practical nonparametric statistics. Wiley (1999)
7. Conover, W.J., Iman, R.L.: Multiple-comparisons procedures. informal report. Tech. rep., Los Alamos National Lab.(LANL), Los Alamos, NM (United States) (1979)
8. Fedorov, A., Beichel, R., Kalpathy-Cramer, J., Finet, J., Fillion-Robin, J.C., Pujol, S., Bauer, C., Jennings, D., Fennessy, F., Sonka, M., et al.: 3d slicer as an image computing platform for the quantitative imaging network. Magnetic Resonance Imaging **30**(9), 1323–1341 (2012)
9. Fischler, M.A., Bolles, R.C.: Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Communications of the ACM **24**(6), 381–395 (1981)

10. Frazer, L.L., Louis, N., Zbijewski, W., Vaishnav, J., Clark, K., Nicolella, D.P.: Super-resolution of clinical ct: Revealing microarchitecture in whole bone clinical ct image data. Bone **185**, 117115 (2024)
11. Galassi, A., Martín-Guerrero, J.D., Villamor, E., Monserrat, C., Rupérez, M.J.: Risk Assessment of Hip Fracture Based on Machine Learning. Applied Bionics and Biomechanics **2020**, 8880786 (2020)
12. Hildebrand, T., Rüegsegger, P.: A new method for the model-independent assessment of thickness in three-dimensional images. Journal of microscopy **185**(1), 67–75 (1997)
13. Isensee, F., Jaeger, P.F., Kohl, S.A., Petersen, J., Maier-Hein, K.H.: nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. Nature methods **18**(2), 203–211 (2021)
14. Johnson, H., Harris, G., Williams, K., et al.: Brainsfit: mutual information rigid registrations of whole-brain 3d images, using the insight toolkit. Insight J **57**(1), 1–10 (2007)
15. Kanis, J.A., Norton, N., Harvey, N.C., Jacobson, T., Johansson, H., Lorentzon, M., McCloskey, E.V., Willers, C., Borgström, F.: SCOPE 2021: a new scorecard for osteoporosis in Europe. Archives of Osteoporosis **16**(1), 82 (2021)
16. Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z., 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. pp. 4681–4690 (2017)
17. Lepcha, D.C., Goyal, B., Dogra, A., Goyal, V.: Image super-resolution: A comprehensive review, recent trends, challenges and applications. Information Fusion **91**, 230–260 (2023)
18. Maquer, G., Musy, S.N., Wandel, J., Gross, T., Zysset, P.K.: Bone Volume Fraction and Fabric Anisotropy Are Better Determinants of Trabecular Bone Stiffness Than Other Morphological Variables. Journal of Bone and Mineral Research **30**(6), 1000–1008 (2015)
19. Mc Donnell, P., Mc Hugh, P.E., O' Mahoney, D.: Vertebral Osteoporosis and Trabecular Bone Quality. Annals of Biomedical Engineering **35**(2), 170–189 (2007)
20. Nishiyama, K.K., Shane, E.: Clinical imaging of bone microarchitecture with HR-pQCT. Current Osteoporosis Reports **11**(2), 147–155 (2013)
21. Porter, J.L., Varacallo, M.A.: Osteoporosis. In: StatPearls. StatPearls Publishing, Treasure Island (FL) (2025), http://www.ncbi.nlm.nih.gov/books/NBK441901/
22. Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B.: High-resolution image synthesis with latent diffusion models. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 10684–10695 (2022)
23. Wang, X., Xie, L., Dong, C., Shan, Y.: Real-esrgan: Training real-world blind super-resolution with pure synthetic data. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 1905–1914 (2021)
24. Wang, X., Yu, K., Wu, S., Gu, J., Liu, Y., Dong, C., Qiao, Y., Change Loy, C.: Esrgan: Enhanced super-resolution generative adversarial networks. In: Proceedings of the European conference on computer vision (ECCV) workshops. pp. 0–0 (2018)
25. Wang, Z.: Image quality assessment: from error visibility to structural similarity. IEEE transactions on image processing **13**(4), 600–612 (2004)
26. Wasserthal, J., Breit, H.C., Meyer, M.T., Pradella, M., Hinck, D., Sauter, A.W., Heye, T., Boll, D.T., Cyriac, J., Yang, S., et al.: Totalsegmentator: robust segmentation of 104 anatomic structures in ct images. Radiology: Artificial Intelligence **5**(5) (2023)

27. Yue, Z., Wang, J., Loy, C.C.: Resshift: Efficient diffusion model for image super-resolution by residual shifting. Advances in Neural Information Processing Systems **36** (2024)
28. Zhang, R., Isola, P., Efros, A.A., Shechtman, E., Wang, O.: The unreasonable effectiveness of deep features as a perceptual metric. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 586–595 (2018)
29. Zhou, Q.Y., Park, J., Koltun, V.: Open3D: A modern library for 3D data processing. arXiv:1801.09847 (2018)