Why do State-of-the-art Super-Resolution Methods not work well for Bone Microstructure CT Imaging?

Rehan Jhuboo *, Ievgen Redko *, Alain Guignandon †, Françoise Peyrin ‡, and Marc Sebban *
*Laboratoire Hubert Curien, Saint-Etienne, France, firstname.lastname@univ-st-etienne.fr
† SAINBIOSE, Saint-Etienne, France, alain.guignandon@univ-st-etienne.fr
‡ Univ Lyon, CNRS 5220, INSERM 1294, Creatis, INSA Lyon, France, francoise.peyrin@creatis.insa-lyon.fr

***Abstract*—3D Computerized Tomography (CT) is a gold standard technique to assess bone microstructure in the context of bone diseases such as osteoporosis. However, when acquired *in-vivo*, bone images may suffer from a low spatial resolution and the presence of noise due to the limited tolerable radiation exposure. One way to overcome this issue consists in applying Super-Resolution (SR) techniques that aim at recovering high resolution images. Significant progress has been recently made thanks to deep learning SR methods trained on natural image datasets. To measure the reconstruction quality, Peak Signal to Noise Ratio (PSNR) and Structural Similarity (SSIM) are commonly used in the SR literature. In this paper, we give evidence of the limitation of these two criteria. Through extensive experiments performed from a dataset of mice tibias specifically collected and imaged for this study, we show that state of the art deep learning-based SR methods miss important details about the bone microstructure which is not reflected by the PSNR and SSIM values. This study opens the door to future promising lines of research including new SR methods regularized with respect to morphometric and topological parameters of bone microstructures.**

***Index Terms*—super resolution, bones, CT-images**

I. Introduction

Computerized Tomography (CT) images are fundamental to assess bone diseases such as osteoporosis. High-Resolution peripheral Quantitative Computed Tomography (HR-pQCT) can provide images of bone microstructures in humans *in-vivo* at spatial resolution of the order of 100µm. More resolved techniques (*e.g.* micro-CT, nano-CT) can be used to better understand bone diseases but they are restricted to small animals or humans *ex-vivo*. Recovering High Resolution (HR) images from their Low Resolution (LR) counterparts is a challenging task, called Super Resolution (SR), that received much attention from the signal/image processing and machine learning communities. In bone imaging, SR methods are a real opportunity to overcome the limited tolerable radiation exposure of humans. For instance, by taking HR-pQCT images of astronauts, they might offer a solution for observing new bone formation or resorption as well as porosity evolution after repeated space flights.

Deep learning (DL)-based SR methods allowed to obtain quite remarkable results during the past few years. They typically resort to a combination of convolutional layers before an upscaling step to output the SR image (*e.g.* transposed convolutional layer, subpixel convolution layer, etc.). The first DL-based SR method in the form of a convolutional network, called SRCNN, was introduced in [1] and improved by FSRCNN [2] to overcome its high computational cost. Other DL SR architectures have been proposed in the literature, including SRResNet [3] which makes use of a residual network with skip connections between layers, or Generative Adversarial Networks (GAN) [4] as introduced in SRGAN [3] and enhanced with ESRGAN [5]. On the other hand, RCAN [6] introduced an attention layer to the architecture, becoming the new state of the art of the so-called non blind SR methods, the family of networks assuming a known degradation (typically bicubic). More recently, blind methods have been introduced aiming at estimating a degradation kernel used for restoring SR images. Among them, DAN [7] is probably the most performing method nowadays. Note that DL-based SR has received a growing interest in the medical field, mostly on MRI images (e.g. [8], [9]) and a few focusing on CT images (e.g. [10], [11] and [12]). Those approaches are usually based on the aforementioned SR methods.

To measure the reconstruction quality of the SR images, two main criteria are commonly used: Peak Signal to Noise Ratio (PSNR) and Structural SIMilarity index (SSIM). The former aims at quantifying the reconstruction quality of images (or videos) subject to lossy compression. The latter [13] measures the similarity between the SR image and the ground truth (HR). While these two measures are indisputably relevant for estimating from an aesthetic perspective the perceptual quality of SR images representing animals, flowers, human faces, buildings or landscapes, we claim that they are not well adapted for dealing with bone CT images.

The contribution of this paper is two-fold: first, we perform an extensive experimental comparison between four state of the art DL-based SR (blind and non blind) methods on a dataset of mice tibias specifically collected and imaged for this study. We show that while both PSNR and SSIM reach reasonable scores for some methods giving evidence of the good perceptual quality of the restored images, this is at the expense of missing important details about the bone microstucture that is essential for establishing clinical diagnosis; second, we highlight the importance of some morphometric and topological parameters [14] characterizing the trabecular and cortical bones and show that the studied state of the

This work was funded by EUR MANUTECH SLEIGHT (N°ANR-16-IDEX-0005) and was performed within the framework of the LABEX PRIMES (ANR-11-LABX-0063) of Université de Lyon, within the program "Investissements d'Avenir" (ANR).

art SR methods do not lead to any enhancement of these biological parameters compared to the LR images. This study should encourage the next generation of SR methods to take advantage of these parameters to design new loss functions or regularization terms based on this biological knowledge. However, assessing quantitatively the microstructure of a bone is challenging because this latter has a hierarchical organization including structures at different scales. Indeed, trabecular bone micro architectures are made of a complex network of small trabeculae while the cortical bone constitutes the external layer of the bone and includes pores at various scales.

The paper is organized as follows: Section II is devoted to preliminary information about SR and the presentation of the related work; Section III focuses on evaluation metrics and presents morphometric and topological parameters that can be used to evaluate the quality of SR bone images; Section IV is dedicated to a comparative experimental study between SR methods. Finally, we conclude in Section V opening new promising lines of research.

II. Preliminaries and Related work

We can formalize the super resolution (SR) problem as follows. Let $\mathbf{f}$ and $\mathbf{g}$ be respectively the high resolution (HR) and low resolution (LR) images. We assume that there exists an underlying degradation function D such that: $\mathbf{g} = D(\mathbf{f})$. The SR task consists in restoring $\mathbf{f}$ from $\mathbf{g}$. The output of a SR convolutional neural network S takes the following form: $\hat{\mathbf{f}} = S(\mathbf{g}, \theta)$, where $\hat{\mathbf{f}}$ is the estimated HR image and θ are the parameters of the neural network. S is trained using $(\mathbf{g}, \mathbf{f})$-pairs of LR and HR images and minimizing a loss function often based on the mean squared reconstruction error (MSE).

The first DL-based SR method, called SRCNN, was presented by Dong et al. [1]. The corresponding architecture is composed of three convolutional layers allowing to perform three successive operations from LR images: (i) the extraction of a set of feature maps, (ii) a non linear mapping of these features followed by (iii) the reconstruction of the HR image. To overcome the high computational cost of SRCNN, an improved version called FSRCNN [2] takes advantage of a deconvolution layer at the end of the network allowing the mapping to be learned directly from the LR images. Ledig et al. [3] designed the first SR method based on a generative adversarial network (GAN), called SRGAN. Unlike SRCNN and FSRCNN, SRGAN does not minimize the mean square reconstruction error. It rather resorts to a perceptual loss function composed of an adversarial loss and a content loss, which aim at inferring more photo-realistic images. To get rid of some unpleasant artifacts, ESRGAN [5] modifies the architecture and the two losses of SRGAN leading to an enhancement in terms of sharpness and details. On the other hand, Zhang et al. introduced RCAN [6], a residual channel attention network composed of several residual groups with long skip connections allowing the network to better focus on high-frequency information.

Acquiring a large number of $(\mathbf{g}, \mathbf{f})$-pairs for training the previous DL networks can be a tricky task according to the application at hand. A cheaper and common way to create the training set consists in generating artificially the LR images according to an priori known degradation function D. SR methods that are following this strategy are usually called *non blind*. SRCNN, FSRCNN, RCAN and SRGAN belong to this family because they all resort to a bicubic downsampling. Because of the limitation of this model that can be seen as a too constrained prior inducing a domain gap between bicubically generated LR examples and real images, a new family of *blind* models has recently emerged [15] supposing that the degradation is unknown and has to be estimated through a kernel and an additive noise.

DAN [7] is a well known representative of this family. It performs an alternating optimization process allowing a joint estimation of the blur kernel and the restoration of the SR images. On the other hand, Kernel-GAN [16] estimates the degradation kernel based on internal statistics. It learns only from the LR test image at test time and optimizes its internal distribution of patches. This latter has been shown to play a key role to identify the right SR kernel. However, it suffers from a time consuming training phase, which is an important drawback for 3D images containing hundreds of slices. To finish this (non exhaustive) list of related work, note that the algorithm pix2pix [17], which is not a specific SR method, is very known in medical image processing for image segmentation/reconstruction. It is based on a GAN architecture with a U-net generator, composed of an encoder-decoder generator, and a patch-GAN discriminator. Since this algorithm takes the form of an image-to-image translation, it can be used - provided some slight changes (as discussed in the experimental part) - to perform an SR task.

III. SR evaluation metrics and bone biological parameters

The peak signal to noise ratio (PSNR) and the Structural SIMilarity index (SSIM) are two measures that are the most widely used to evaluate SR methods. PSNR quantifies the reconstruction quality of images subject to lossy compression. It corresponds to a function of the ratio of (i) the maximum possible value of a pixel (*i.e.* 255 for 8-bit images) over (ii) the mean square error between the restored image (SR) and the ground truth (HR). Therefore, the higher the PSNR, the better the SR is. SSIM [13] is a measure of the perceptual quality of the reconstruction. It compares the SR and HR images through a weighted combination of three comparison measurements: the structure, the luminance and the contrast. SSIM is symmetric and ranges between [0,1]. It is equal to 0 if there is no structural similarity and 1 if the images are identical.

Even though PSNR and SSIM are widely used in the SR field, we claim here that those two metrics are not well adapted to address bone SR tasks. To give evidence of the issues raised by these two criteria, we report in Fig. 1 an illustration of the resulting images (representing mice tibias) obtained by four state of the art SR methods (RCAN, FSRCNN, pix2pix and DAN) as well as a bicubic upscaling obtained from a real

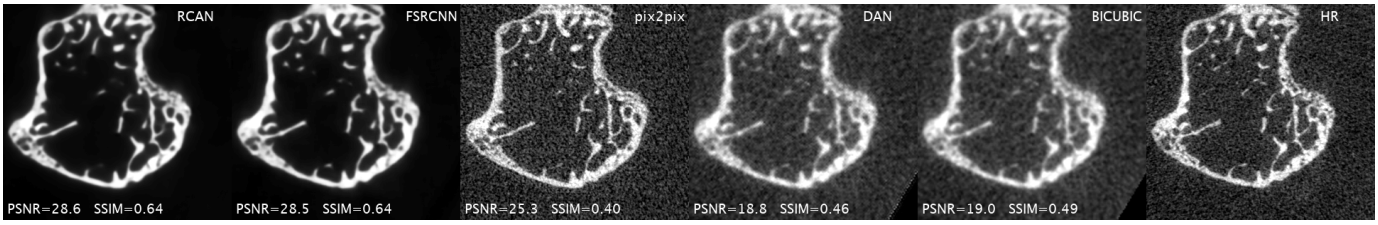

Fig. 1. SR images obtained from RCAN, FSRCNN, pix2pix and DAN. A bicubic upscaling (on the left) is shown as a baseline. For each method, the PSNR and SSIM values are reported. The ground truth is given by the HR image.

LR image, used as a baseline. We can note that RCAN and FSRCNN are the best methods from a perceptual standpoint, illustrated by the highest PSNR and SSIM scores. But the latter are not as high as the values obtained in the literature on standard datasets shared by the SR community. Moreover, pix2pix and DAN provide pretty poor results (even worse for the latter compared to the bicubic upscaling) illustrating the fact than bone SR is a complex task. The most important remark we can make from this figure is that even though from an aesthetic perspective RCAN and FSRCNN seem to be pretty convincing, we can observe from the HR image that details such as pores of the cortical bone and some parts of the trabecular bone are lost by these SR methods, while they play a key role in the medical diagnosis of bone diseases.

From a biological perspective, several 3D morphometric and topological parameters characterize well the bone microstructure and should be used to evaluate SR methods. In the experiments, we will use the following trabecular parameters: **Bone Volume/Total Volume (BV/TV)**, the percentage of bone in the Region of Interest (ROI); **Trabecular thickness**, the average thickness of trabeculae; **Conn.D**, the connectivity density; **Trabecular Separation**, the average distance between trabeculae; **Degree of Anisotropy**, indicator of how strongly oriented are trabeculae. The main cortical parameters are the following: **close and open porosity, number of pores or cortical thickness**. Due to lack of space, we will not show results on those parameters, but note that the main conclusions made on the trabecular parameters hold for the cortical ones.

Note that more human perception correlated evaluation metrics exist [18]. Nevertheless, PSNR and SSIM are more frequently used than those metrics in papers dealing with SR.

IV. Experiments

We present here the dataset that has been specifically collected and imaged for this study. Then we report the results and main conclusions obtained from an extensive experimental comparison between four state of the art SR methods.

A. Dataset and experimental setup

Forty mice tibias were imaged from a Scanco micro-CT (μCT) at the laboratory Sainbiose, Saint-Etienne. 3D LR and HR images with cubic voxel size of 19 μm and 10.5 μm respectively were acquired on the same tibia site for all mice. We thus created a 40 mice dataset of LR and HR 3D images. The LR and HR slices were not directly paired since the HR (resp. LR) slices are spaced by 10.5 μm (resp. 19 μm) and the tibias may have been scanned with a different orientation and position for the two resolutions. Thus, we first performed 3D image registration in order to create paired (HR,LR) slices to be used for training and testing. For registration, we used an affine transform from the AntsPy[1] package in python. An interpolation was applied to the 3D LR images before registration in order to manipulate similar bone volumes in both images. After registration, we obtained 26484 HR slices[2].

In this series of experiments, we compared four SR methods: RCAN, FSRCNN, DAN and pix2pix. For the first three, we used the versions available online. For the latter, we used the implementation of You et al. [19] where the usual noise input of GAN were replaced by our LR images, and the loss function was changed accordingly. Each SR method was trained from scratch from 29 (LR,HR) pairs of our dataset. This training set was separated into $90\%/10\%$ for the learning/validation process for DAN and RCAN. For pix2pix and FSRCNN, we used hyperparameters obtained from a preliminary study (not reported in this paper). We used 11 bone images for testing and calculated SSIM and PSNR from the resulting 6948 (SR,HR) pairs of slices. As a baseline, we also used a bicubic upscaling of LR images which is supposed to be the worst SR result.

The biological bone parameters were calculated from the 3D reconstructions from the SR images ouput from the different SR methods. The computation of these 3D parameters was performed after the binarization of the images in bone (white) and background (black) by using the 2-class Otsu method. Since some HR images can be subject to the presence of noise, we first applied a total variation denoising step to the HR images before segmentation. We also achieved this task on the SR images generated by DAN and pix2pix which do not benefit from the effect of the MSE loss function, like RCAN and FSRCNN. As indicated before, we only present in this section the 3D trabecular bone parameters. This required to select a volume of interest of 50 consecutive slices where trabecular bone is mainly present. Then, we created a binary mask of the trabecular area to exclude the cortical bone. The computation of parameters was performed by using the Skyscan CTan software.

[1]https://antspyx.readthedocs.io/en/latest/

[2]The source code and the datasets used in this paper are available at https://github.com/RJhuboo/SRBoneMicrostructure

TABLE I

MORPHOMETRIC AND TOPOLOGICAL PARAMETERS EVALUATION OF THE 4 SUPER-RESOLUTION METHODS. RESULTS FOR THE GROUND TRUTH HR AND THE ACTUAL LR ARE ALSO REPORTED. THE RESULTS ARE REPORTED IN TERMS OF MEANS AND STANDARD DEVIATIONS CALCULATED FROM THE ELEVEN TEST 3D VOLUMES OF INTEREST. **BOLD** VALUES ARE THE BEST SR RESULTS.

Parameter	Super-Resolution Method					
	HR	***RCAN***	***FSRCNN***	***pix2pix***	***DAN***	***LR***
Trabecular thickness (µm)	61.7 ± 7.8	69.6 ± 9.4	69.6 ± 9.0	$\mathbf{59.5 \pm 6.9}$	77.9 ± 12.8	67.6 ± 10.3
BV/TV (%)	9.8 ± 3.5	8.6 ± 3.8	$\mathbf{9.5 \pm 3.4}$	7.9 ± 2.9	11.7 ± 5.3	10.0 ± 4.1
Trabecular separation (µm)	291 ± 31.9	366 ± 42.7	331 ± 33.2	330 ± 41.7	$\mathbf{321 \pm 50.2}$	331 ± 36.4
Degree of anisotropy	1.7 ± 0.4	2.1 ± 0.8	1.8 ± 0.5	1.8 ± 0.6	$\mathbf{1.7 \pm 0.4}$	2.0 ± 0.4
Conn.D (mm^{-3})	30 ± 18.6	$\mathbf{31.8 \pm 13.4}$	40 ± 16.5	18.2 ± 1.10	20 ± 12.1	48.2 ± 23.7

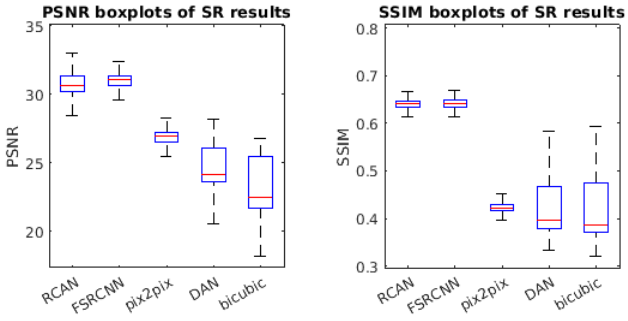

Fig. 2. PSNR comparison of four SR methods as well as a bicubic upscaling as a baseline. Red line: median. Blue box: 15^{th} and 75^{th} percentiles. Black lines: upper and lower adjacent values.

B. Results

Fig. 2 displays boxplots of the PSNR and SSIM for the different SR methods. While RCAN and FSRCNN allow to obtain reasonable scores (PSNR greater than 30 and SSIM around 0.65), we can note that they do not reach values usually obtained in the literature on standard datasets composed of natural images. These results illustrate well the fact that bone SR is a much more challenging task than restoring images of buildings, flowers or animals. Even worse, we can note that DAN, known to be the state of the art blind SR method, is only slightly better than the bicubic baseline that shows its difficulty to capture the right degradation. We report in Table I the values of some trabecular bone parameters. It is worth noticing that the best two methods RCAN and FSRCNN in terms of PSNR and SSIM fail to recover well the bone microstructure. Both tend to produce denoised and blurry SR images (as illustrated in Fig. 1) where small trabeculae disappeared leading to an overestimation of the trabecular separation and thickness (compared to the ground truth HR) and thus an underestimation of BV/TV. On the other hand, pix2pix produces SR images with more noise that prevents the recovery of the right connectivity. DAN tends to overestimate the trabecular thickness (as well illustrated in Fig. 1). Associated with the noise it generates, DAN also looses information about the connectivity density. Finally, the most striking result from Table I comes from the fact that the SR parameters are on average worst than those calculated from the LR images. This means that even if the perceptual aspect seems to be improved (especially with RCAN and FSRCNN as expressed by reasonable PSNR and SSIM), the SR methods do not allow any enhancement from a biological perspective and thus are not really efficient for medical diagnosis.

To illustrate graphically how the SR methods over/under-estimate the biological parameters, Fig. 4 shows three SR bone parameters versus their HR counterparts, where the bisecting line $y = x$ corresponds to the perfect matching with the ground truth. This figure confirms that no method is able to recover correctly the bone microstructure w.r.t all parameters.

V. CONCLUSION AND PERSPECTIVES

In this paper, we analyzed the behavior of DL-based SR methods when addressing a super resolution task from micro-CT images of bones. As far as we know, it is the first study showing that state of the art SR methods (i) lead to PSNR and SSIM values that are reasonable but smaller than what is generally observed on standard datasets of natural images and (ii) do not enhance LR images when evaluating the capacity of the methods to recover the bone microstructure. Therefore, this study raises the question of the direct use of these methods for medical diagnosis. It also highlights the fundamental difference between inducing aesthetic images and generating SR images relevant from a biological perspective.

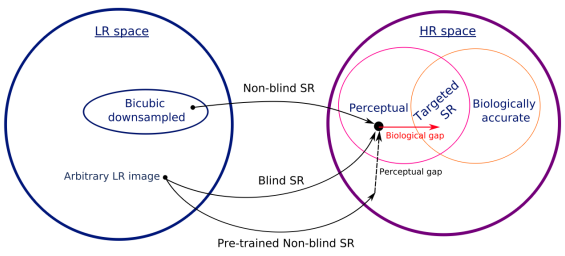

Fig. 3. Illustration of the biological gap existing between images produced by SR methods and images that would satisfy some biological requirements.

This paper opens promising future lines of research. In particular, it raises the need of a new generation of SR methods that would bridge the gap between *perceptual* and *biologically accurate* SR images. Inspired by [15] (Fig. 2), we illustrate this gap by the red arrow depicted in Fig.3. State of the art SR methods generate perceptual images that are relevant when dealing with applications where the easthetics is key. But this is not enough in biological tasks. One way to

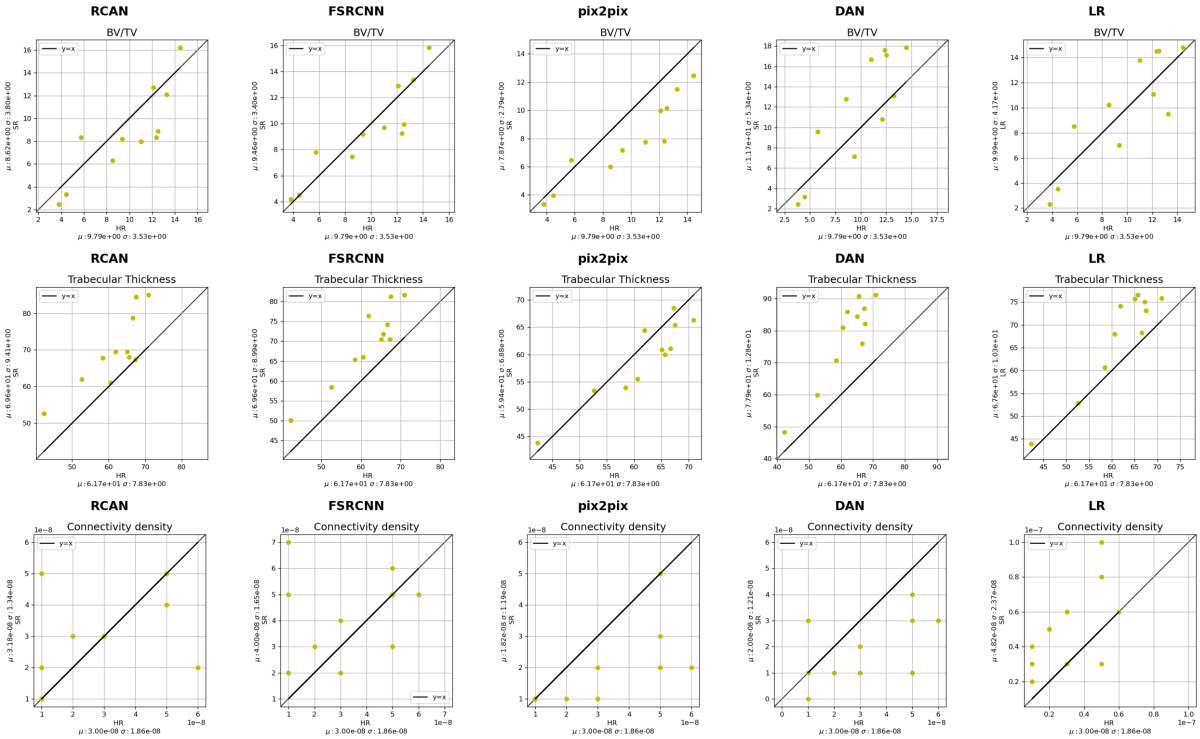

Fig. 4. Comparison of the BV/TV, trabecular thickness and Connectivity density of SR images versus the ground truth HR. Each point above (resp. below) the bisecting line $y = x$ expresses an over (resp. under)-estimation of the parameters by the corresponding SR method. Each point represents one of the 11 mouses used at test time.

address this limitation would consist in pre-training a module calculating the morphometric and topological parameters of bone microstructures and regularizing the loss function so as to reconstruct SR images that are biologically accurate. A more promising and complex strategy would be to jointly learn the super resolution and this biological module. The difficulty coming from the need to build a differentiable module.

REFERENCES

[1] C. Dong, C. C. Loy, K. He, and X. Tang, "Image Super-Resolution Using Deep Convolutional Networks," *arXiv:1501.00092*, 2015.

[2] C. Dong, C. C. Loy, and X. Tang, "Accelerating the Super-Resolution Convolutional Neural Network," *arXiv:1608.00367*, 2016.

[3] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, "Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network," *arXiv:1609.04802*, 2017.

[4] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative Adversarial Networks," *arXiv:1406.2661*, 2014.

[5] X. Wang, K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, Y. Qiao, and C. C. Loy, "ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks," *Computer Vision – ECCV*, 2018.

[6] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu, "Image Super-Resolution Using Very Deep Residual Channel Attention Networks," *arXiv:1807.02758*, 2018.

[7] Z. Luo, Y. Huang, S. Li, L. Wang, and T. Tan, "Unfolding the Alternating Optimization for Blind Super Resolution," *arXiv:2010.02631*, 2020.

[8] I. Sanchez and V. Vilaplana, "Brain MRI super-resolution using 3D generative adversarial networks," *arXiv:1812.11440*, 2018.

[9] Y. Chen, F. Shi, A. G. Christodoulou, Z. Zhou, Y. Xie, and D. Li, "Efficient and Accurate MRI Super-Resolution using a Generative Adversarial Network and 3D Multi-Level Densely Connected Network," *arXiv:1803.01417*, 2018.

[10] M.-I. Georgescu, R. T. Ionescu, and N. Verga, "Convolutional Neural Networks with Intermediate Loss for 3D Super-Resolution of CT and MRI Scans," *IEEE Access*, vol. 8, pp. 49112–49124, 2020.

[11] Y. Wang, Q. Teng, X. He, J. Feng, and T. Zhang, "CT-image Super Resolution Using 3D Convolutional Neural Network," *Computers & Geosciences*, vol. 133, p. 104314, Dec. 2019.

[12] H. Yu, D. Liu, H. Shi, H. Yu, Z. Wang, X. Wang, B. Cross, M. Bramler, and T. S. Huang, "Computed tomography super-resolution using convolutional neural networks," in *IEEE International Conference on Image Processing (ICIP)*, pp. 3944–3948, Sept. 2017. ISSN: 2381-8549.

[13] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image Quality Assessment: From Error Visibility to Structural Similarity," *IEEE Transactions on Image Processing*, vol. 13, pp. 600–612, Apr. 2004.

[14] D. W. Dempster, J. E. Compston, M. K. Drezner, F. H. Glorieux, J. A. Kanis, H. Malluche, P. J. Meunier, S. M. Ott, R. R. Recker, and A. M. Parfitt, "Standardized nomenclature, symbols, and units for bone histomorphometry: a 2012 update of the report of the ASBMR Histomorphometry Nomenclature Committee," *Journal of Bone and Mineral Research*, vol. 28, pp. 2–17, Jan. 2013.

[15] A. Liu, Y. Liu, J. Gu, Y. Qiao, and C. Dong, "Blind Image Super-Resolution: A Survey and Beyond," *arXiv:2107.03055*, 2021.

[16] S. Bell-Kligler, A. Shocher, and M. Irani, "Blind Super-Resolution Kernel Estimation using an Internal-GAN," in *Advances in Neural Information Processing Systems*, vol. 32, Curran Associates, Inc., 2019.

[17] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-Image Translation with Conditional Adversarial Networks," *arXiv:1611.07004*, 2018.

[18] H. Chen, X. He, L. Qing, Y. Wu, C. Ren, and C. Zhu, "Real-world single image super-resolution: A brief review," *Elsevier*, 2022.

[19] C. You, G. Li, Y. Zhang, X. Zhang, H. Shan, M. Li, S. Ju, Z. Zhao, Z. Zhang, W. Cong, M. W. Vannier, P. Saha, E. Hoffman, and G. Wang, "CT super-resolution GAN constrained by the identical, residual, and cycle learning ensemble (gan-circle)," *IEEE Transactions on Medical Imaging*, vol. 39, pp. 188–203, 06 2020.