RPLHR-CT Dataset and Transformer Baseline for Volumetric Super-Resolution from CT Scans

Pengxin Yu[1], Haoyue Zhang[1,2], Han Kang[1], Wen Tang[1], Corey W. Arnold[2],
and Rongguo Zhang[1](✉)

[1] Infervision Medical Technology Co., Ltd., Beijing, China
zrongguo@infervision.com
[2] Computational Diagnostics Lab, UCLA, Los Angeles, USA

Abstract. In clinical practice, anisotropic volumetric medical images with low through-plane resolution are commonly used due to short acquisition time and lower storage cost. Nevertheless, the coarse resolution may lead to difficulties in medical diagnosis by either physicians or computer-aided diagnosis algorithms. Deep learning-based volumetric super-resolution (SR) methods are feasible ways to improve resolution, with convolutional neural networks (CNN) at their core. Despite recent progress, these methods are limited by inherent properties of convolution operators, which ignore content relevance and cannot effectively model long-range dependencies. In addition, most of the existing methods use pseudo-paired volumes for training and evaluation, where pseudo low-resolution (LR) volumes are generated by a simple degradation of their high-resolution (HR) counterparts. However, the domain gap between pseudo- and real-LR volumes leads to the poor performance of these methods in practice. In this paper, we build the first public real-paired dataset RPLHR-CT as a benchmark for volumetric SR, and provide baseline results by re-implementing four state-of-the-art CNN-based methods. Considering the inherent shortcoming of CNN, we also propose a transformer volumetric super-resolution network (TVSRN) based on attention mechanisms, dispensing with convolutions entirely. This is the first research to use a pure transformer for CT volumetric SR. The experimental results show that TVSRN significantly outperforms all baselines on both PSNR and SSIM. Moreover, the TVSRN method achieves a better trade-off between the image quality, the number of parameters, and the running time. Data and code are available at https://github.com/smilenaxx/RPLHR-CT.

Keywords: Volumetric super-resolution · CT · Deep learning · Transformer

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/978-3-031-16446-0_33.

L. Wang et al. (Eds.): MICCAI 2022, LNCS 13436, pp. 344–353, 2022.
https://doi.org/10.1007/978-3-031-16446-0_33

1 Introduction

Volumetric medical imaging, such as computed tomography (CT) and magnetic resonance imaging (MRI), is an important tool in diagnostic radiology. Although high-resolution volumetric medical imaging provides more anatomical and functional details that benefit diagnosis [3,22,24], long acquisition time and high storage cost limit the wide application in clinical practice. As a result, it is routine to acquire anisotropic volumes in practice, which have high in-plane resolution and low through-plane resolution. However, the disparity in resolution can lead to several challenges: (1) the inability to display sagittal or coronal views with adequate detail [10]; (2) the insufficiency of spatial resolution to observe the details of lesions [24] and; (3) the challenge to the robustness of 3D medical image processing algorithms [8,16]. A feasible solution is to use super-resolution (SR) algorithms [26] to upsample anisotropic volumes along the depth dimension, in order to restore high resolution (HR) from low resolution (LR). This approach is referred to as “volumetric SR.”

CNN-based algorithms have achieved outstanding performance in SR for natural images [20] and these techniques have been introduced for volumetric SR [1,4,6,12,13,15,17,18,23,25]. Though significant advances have been made, CNN-based algorithms remain limited by the inherent weaknesses of convolution operators. On the one hand, using the same convolution kernel to restore various regions may neglect the content relevance. Liu et al. [13] take this into consideration and propose a multi-stream architecture based on lung segmentation to recover different regions separately, but this is hard to be a one-size-fits-all solution. On the other hand, the non-local content similarity of images has been used as an effective prior in image restoration [27]. Unfortunately, the local processing principle of the convolution operator makes algorithms difficult to effectively model long-range dependence. Recently, transformer networks have shown good performance in several visual problems of natural image [5,14], including SR [2,11]. Self-attention mechanism is the key to the success of transformer. Compared to CNN-based algorithms, transformer can model long-range dependence in the input domain and perform dynamic weight aggregation of features to obtain input-specific feature representation enhancement [9]. These results prompted us to explore a transformer-based SR method.

Another impediment to the application of volumetric SR methods is data. Most relevant studies use HR volume as ground truth and degrade it to construct paired pseudo-LR volumes with which to train and evaluate methods [4,15,17,18,23,25]. For instance, Peng et al. [17] perform sparse sampling on the depth dimension of thin CT to obtain pseudo thick CT. Zhao et al. [25] simulate pseudo-LR MRI by applying an ideal low-pass filter to the isotropic T2-weighted MRI followed by an anti-ringing Fermi filter. However, the performance will be affected when test on the real-LR volume [1] because of the domain gap between pseudo- and real-LR volume. To avoid it, some studies collect real-paired LR-HR volumes [1,6,12,13,16]. For example, Liu et al. [13] collect 880 real pairs of chest CTs and construct a progressive upsampling model to reconstruct 1 mm CT from 5 mm CT. In the field of MRI, a large data set containing 1,611 real pairs of T1-weighted MRIs have been used to develop the proposed SCSRN method [12]. However, a benchmark to objectively evaluate various volumetric SR methods is still lacking.

To address this deficiency, the first goal of this work is to curate a medium-sized dataset, named Real-Paired Low- and High-Resolution CT (RPLHR-CT), for volumetric SR. RPLHR-CT contains real-paired thin-CTs (slice thickness 1 mm) and thick-CTs (slice thickness 5 mm) of 250 patients. To the best of our knowledge, RPLHR-CT is the first benchmark for volumetric SR, which enables method comparison. The other goal of our work is to explore the potential of transformer for volumetric SR. Specifically, we propose a novel Transformer Volumetric Super-Resolution Network (TVSRN). TVSRN is designed as an asymmetric encoder-decoder architecture with transformer layer, without any convolution operations. TVSRN is the first pure transformer used for CT volumetric SR. We re-implement and benchmark state-of-the-art CNN-based volumetric SR algorithms developed for CT and show that our TVSRN outperforms existing algorithms significantly. Additionally, TVSRN achieves a better trade-off between image quality, the number of parameters, and running time.

2 Dataset and Methodology

2.1 RPLHR-CT Dataset

Dataset Description. The RPLHR-CT dataset is composed of 250 paired chest CTs from patients. All data have been anonymized to ensure privacy. Philips machines were used to perform CT scans and the raw data were then reconstructed to thin CT (1 mm) and thick CT (5 mm) images. Thus, recovering thin CT (HR volume) from thick CT (LR volume) for this dataset is a volumetric SR task with an upsampling factor of 5 in the depth dimension. The CT scans are saved in NIFTI (.nii) format with volume sizes of $L \times 512 \times 512$, where 512×512 is the size of CT slices, and L is the number of CT slices, ranging from 191 to 396 for thin CT and 39 to 80 for thick CT. The thin CT and the corresponding thick CT have the same in-plane resolution, ranging in $[0.604, 0.795]$, and are aligned according to spatial location.

Dataset split and Evaluation Metric. We randomly split the RPLHR-CT dataset into 100 train, 50 validation and 100 test CT pairs. For evaluation, we quantitatively assess the performance of all methods in terms of peak signal to noise ratio (PSNR) and structural similarity (SSIM) [21]. Significance is tested by one-sided Wilcoxon signed-rank test.

Dataset Analysis. To analysis the difference between the thin CT and thick CT, we group slices in thin CT and thick CT into three categories of slice-pairs according to their spatial relationship, as shown on the left side of Fig. 1. We use PSNR and SSIM to access the changes in the similarity of three slice-pairs in train, validation and test CT pairs. As shown on the right side of Fig. 1, the results indicate that the similarity of slice-pairs at the same spatial location in thin CT and thick CT, namely **Match**, is the highest, while the similarity decreases significantly as the spatial distance becomes larger.

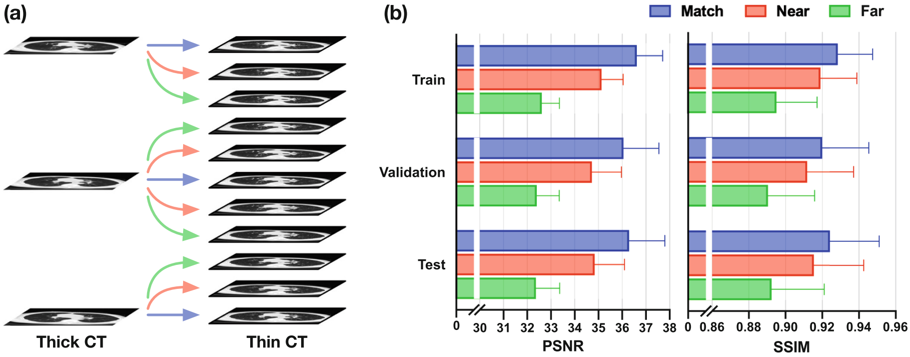

Fig. 1. (a) Three categories of slice-pairs according to their spatial relationship in thin CT and thick CT. Match: same position, shown in blue; Near: 1 mm apart, shown in red; Far: 2 mm apart, show in green. (b) The degree of similarity between the three slice-pairs on the three datasets. (Color figure online)

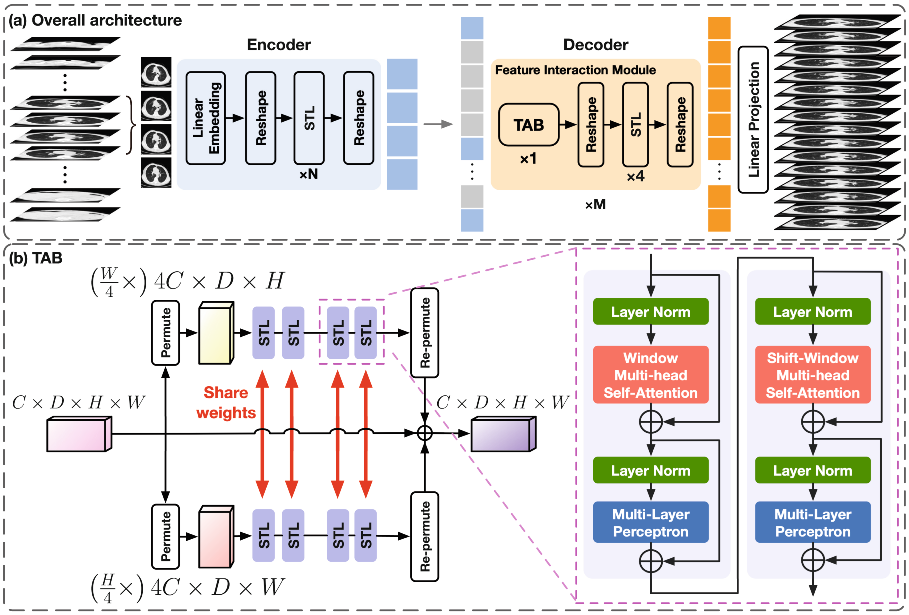

Fig. 2. (a) Illustration of the proposed Transformer Volumetric Super-Resolution Network architecture. (b) Details of TAB. The purple dashed box represents two consecutive swin transformer layers. The batch dimension is indicated in parentheses. (Color figure online)

2.2 Network Architecture

Inspired by MAE [7], we treat volumetric SR as a task to recover the masked regions from the visible regions, where the visible regions refer to the slices in the LR volume and the masked regions refer to the slices in the corresponding HR volume. As illustrated in Fig. 2, we also design our TVSRN with an asymmetric encoder-decoder architecture, but with several targeted modifications. First, in TVSRN, the encoder and the decoder are equally important, and to better model the relationship between the visible regions and the masked regions, the decoder uses a larger amount of parameters than the encoder. Second, instead of the standard transformer layer [5], we use the swin transformer layer (STL) [14], which is less computationally intensive and more suitable for high resolution image, as the basic component of TVSRN. Third, we propose Through-plane Attention Blocks to exploit the spatial positional relationship of volumetric data to achieve better performance.

Encoder is used to map the LR volume to a latent representation. The consecutive slices from LR volumes are denoted as the input $X_e^{in} \in \mathbb{R}^{1\times D\times H\times W}$ of *encoder*, where D, H and W are the depth, height and width, and the channel is 1. X_e^{in} is firstly fed into the *Linear Embedding*, whose number of feature channel is C, to extract shallow features and output $F_s \in \mathbb{R}^{C\times D\times H\times W}$. Then, F_s is reshaped to $F_0 \in \mathbb{R}^{CD\times H\times W}$. We stack N STLs to extract deep features from F_0 as:

$$F_i = H_i^{STL}(F_{i-1}), \quad i = 1, 2, ..., N \tag{1}$$

where $H_i^{STL}(\cdot)$ denotes the i-th STL. Finally, F_N is reshaped to 3D output $X_e^{out} \in \mathbb{R}^{C\times D\times H\times W}$.

Decoder is used to recovery the HR volume from the latent representation. As shown in Fig. 2(a), mask tokens are introduced after the *encoder*, and the full set of X_e^{out} and mask tokens is input to the *decoder* as $X_d^{in} \in \mathbb{R}^{C\times D'\times H\times W}$, where D' is the depth of ground truth. The mask tokens are learned vector that indicates the missing slices in the LR volumes compared to the HR counterpart. *Decoder* stack M Feature Interaction Modules (FIMs), which consists of one Through-plane Attention Block (TAB), four STLs and two reshape operations. The reshape operations are used to reshape the input feature map into the size expected by the next block. The output of the *decoder* is X_d^{out} with the same size as X_d^{in}. Note that the design of asymmetric *decoder* can easily be adapted to other upsampling rates by changing the number of mask tokens.

The details of TAB are illustrated in Fig. 2(b). TAB is the first block in each FIM. There are two parallel branches in TAB that perform self-attention on the input from coronal and sagittal views, respectively. In both views, the depth dimension will become an axis of the STL's window, so the relative position relationship between slices will be incorporated into the calculation. The parameter weights of the corresponding STL on the two parallel branches are shared. Given the input feature z_{in} of TAB, the output is computed as:

$$
\begin{aligned}
z_0^{sag} = P^{sag}(z_{in}),\ z_0^{cor} = P^{cor}(z_{in}) \\
z_j^{sag} = H_j^{STL}(z_{j-1}^{sag}),\ z_j^{cor} = H_j^{STL}(z_{j-1}^{cor}),\ \ j = 1,2,3,4 \\
z_{out} = z_{in} + P_{re}^{sag}(z_4^{sag}) + P_{re}^{cor}(z_4^{cor})
\end{aligned} \tag{2}
$$

where $P^{sag}(\cdot)$ and $P^{cor}(\cdot)$ are permutation operations that transform the input to sagittal and coronal view, respectively. $P_{re}^{sag}(\cdot)$ and $P_{re}^{cor}(\cdot)$ denote re-permutation operations that reshape the input back to original size. In addition, TAB contains residual connection, which allow the aggregation of different levels of features.

Reconstruction Target. The X_d^{out} is fed into the *Linear Projection* to obtain the pixel-wise prediction $\hat{Y} \in \mathbb{R}^{D' \times H \times W}$. The L_1 pixel loss is formulated as:

$$L_{pixel} = \frac{1}{D' \times H \times W} \sum_{k,i,j} |\hat{Y}_{k,i,j} - Y_{k,i,j}| \tag{3}$$

where Y is the ground truth HR volume.

Architecture Hyper-parameters. For each STL, the patch size is 1×1 and the window sizes of x-axis, y-axis and z-axis are set to 8, 8 and 4. For *Linear Embedding*, the channel number C is 8. The number of STLs in encoder and FIMs in decoder is set to $N = 4$ and $M = 1$, respectively.

3 Experiments and Results

Implementation Details. We normalize the intensity of the CT images from $[-1024, 2048]$ to $[0, 1]$. During training, $4 \times 256 \times 256$ cubes from thick CTs are used as input and the corresponding $16 \times 256 \times 256$ cubes from thin CTs are used as ground truth, in where $16 = (4 - 1) \times 5 + 1$. During inference, we feed cubes from thick CTs to the model in a sliding window manner, in which the overlap of depth dimension is 1 and the rest is 0. If the depth of untested cubes is less than 4, we feed the last 4 slices into the model. For multiple predictions on the same coordinate, we take the average as the final value. TVSRN is trained with Adam optimizer. The learning rate is 0.0001 and the batch size is 1. For the comparison methods, we follow descriptions provided in the original papers to re-implement the models, as none have public code available. Settings not detailed in the original paper will remain consistent with our work. Data augmentation include random cropping and horizontal flipping. The framework is implemented in PyTorch, and trained on NVIDIA A6000 GPUs.

3.1 Results and Analysis

Figure 3(a) summarizes the quantitative comparisons of our method and other state-of-the-art CT volumetric SR methods: ResVox [6], MPU-Net [13], SAINT [17] and DA-VSR [18]. For ResVox, the noise reduction part is removed. For MPU-Net, we do not use the multi-stream architecture due to the lack of

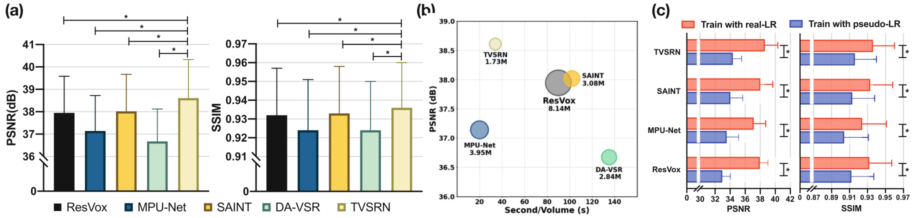

Fig. 3. (a) Quantitative comparisons of our TVSRN and other state-of-the-art methods. $*$ indicates $p < 0.001$. (b) PSNR vs. processing time of each volume with number of parameters shown in circle size. (c) quantitative results of pseudo images experiment.

available lung masks. TVSRN achieves PSNR of 38.609 ± 1.721 and SSIM of 0.936 ± 0.024, outperforms others significantly ($p < 0.001$). Moreover, as shown in Fig. 3(b), compared to other methods, TVSRN achieves a better trade-off in terms of the PSNR (optimal), the number of parameters (optimal), and the running time (suboptimal). We also perform the comparison on an external test set, where TVSRN also achieved the best performance. Detailed numerical results on the internal test set and external test set are presented in the supplementary material. In addition, a sample-by-sample performance scatterplot is given in the supplementary material.

We visualize the axial, coronal and sagittal views of HR CT volume obtained by different methods. It is clear in Fig. 4 that TVSRN has the richest details and the least amount of structural artifacts remaining in different views.

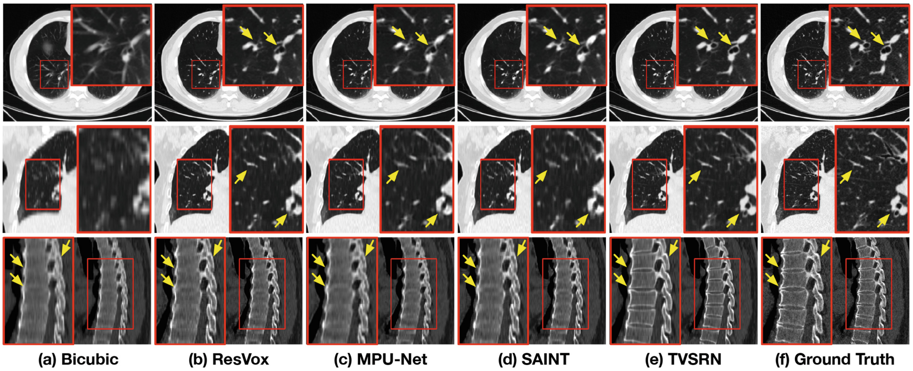

Fig. 4. Visual comparisons of different methods against TVSRN. The first and second rows show the axial view and coronal view respectively, displayed as lung window. The third row is sagittal view, displayed as bone window. Yellow arrows point to areas of marked difference. (Color figure online)

3.2 Domain Gap Analysis

We conduct a pseudo images experiment to illustrate the effect of the domain gap. Specifically, we degrade the training data to obtain pseudo-LR volumes, and use these data to train several different methods. All settings are the same as those in the previous section, except for the training data. For testing, real-LR volumes in the internal test set are used as input to calculate the PSNR and SSIM. As shown in Fig. 3(c), the results show that both PSNR and SSIM of various methods are significantly decreased to varying degrees ($p < 0.001$). Please refer to the supplemental material for more details of degradation.

3.3 Ablation Study

The ablation study is used to verify the contribution of each component in TVSRN on performance. The full TVSRN is compared to:

- TVSRN$_{ViT}^{Encoder}$. A standard transformer-based method based on [5]. We map each patch of size $1 \times 16 \times 16$ to token with length of 512 and set the number of transformer layers to eight. Instead of asymmetric *decoder*, it uses subpixel conversion [19] to perform upsampling.
- TVSRN^{Encoder}. Only the *encoder* of TVSRN was used. N is increasd to eight and C is increased to 32. The upsampling method is subpixel convert.
- TVSRN$^{w/o\ TAB}$. TAB is not used in TVSRN, that is, the relative position relationship among slices is ignored in the network.

Table 1. Results of ablation study for TVSRN in terms of PSNR and SSIM. The best results are **bolded**, and the second best results are underlined. * denotes statistically significant ($p < 0.001$) against above method with one-sided Wilcoxon signed-rank test.

Designs	Param	PSNR(↑)	SSIM(↑)
TVSRN$_{ViT}^{Encoder}$	17.15M	35.537 ± 1.353	0.918 ± 0.026
TVSRN^{Encoder}	1.58M	$38.364 \pm 1.675^*$	$0.934 \pm 0.024^*$
TVSRN$^{w/o\ TAB}$	1.56M	$\underline{38.497 \pm 1.700^*}$	$\underline{0.935 \pm 0.024^*}$
TVSRN	1.73M	$\mathbf{38.609 \pm 1.721^*}$	$\mathbf{0.936 \pm 0.024^*}$

Model performance is summarized in Table 1. Notable observations include: 1) among all designs, TVSRN$_{ViT}^{Encoder}$ has the most parameters but the worst performance, which indicates that it is not feasible to simply apply the transformer to the volumetric SR; 2) replacing standard transformer layer with STL can greatly reduce the number of parameters and improve the performance by a large margin (up to 2.827 dB); 3) asymmetric decoder can improve performance slightly without changing the number of parameters; 4) improvements can be seen from TVSRN$^{w/o\ TAB}$ to TVSRN, indicating the effectiveness of modeling the relative position relationship among slices. A sample-by-sample performance scatterplots in supplemental material is used to further illustrate the effectiveness of individual components.

4 Conclusion

A persistent problem with volumetric SR is the lack of real-paired data for training and evaluation, which makes it challenging generalize algorithms to real-world datasets for practical applications. In this paper, we presented the RPLHR-CT Dataset, which is the first open real-paired dataset for volumetric SR, and provided baseline results by re-implementing four state-of-the-art SR methods. We also proposed a convolution-free transformer-based network, which significantly outperformed existing CNN-based methods and has the least number of parameters and the second shortest running time. In the future, we will enlarge the RPLHR-CT Dataset and investigate new volumetric SR training strategies, such as semi-supervised learning or using unpaired real data.

Acknowledgment. This work was funded by the National Key Research and Development Project (2021YFC2500703).

References

1. Bae, W., Lee, S., Park, G., Park, H., Jung, K.H.: Residual CNN-based image super-resolution for CT slice thickness reduction using paired CT scans: preliminary validation study (2018)
2. Chen, H., et al.: Pre-trained image processing transformer. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 12299–12310 (2021)
3. Chen, M., et al.: Diagnostic performance for erosion detection in sacroiliac joints on MR T1-weighted images: comparison between different slice thicknesses. Eur. J. Radiol. **133**, 109352 (2020)
4. Chen, Y., Shi, F., Christodoulou, A.G., Xie, Y., Zhou, Z., Li, D.: Efficient and accurate MRI super-resolution using a generative adversarial network and 3D multi-level densely connected network. In: Frangi, A.F., Schnabel, J.A., Davatzikos, C., Alberola-López, C., Fichtinger, G. (eds.) MICCAI 2018. LNCS, vol. 11070, pp. 91–99. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00928-1_11
5. Dosovitskiy, A., et al.: An image is worth 16x16 words: transformers for image recognition at scale. In: International Conference on Learning Representations (2020)
6. Ge, R., Yang, G., Xu, C., Chen, Y., Luo, L., Li, S.: Stereo-correlation and noise-distribution aware ResVoxGAN for dense slices reconstruction and noise reduction in thick low-dose CT. In: Shen, D., et al. (eds.) MICCAI 2019. LNCS, vol. 11769, pp. 328–338. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-32226-7_37
7. He, K., Chen, X., Xie, S., Li, Y., Dollár, P., Girshick, R.: Masked autoencoders are scalable vision learners. arXiv preprint arXiv:2111.06377 (2021)
8. 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. Nat. Methods **18**(2), 203–211 (2021)
9. Khan, S., Naseer, M., Hayat, M., Zamir, S.W., Khan, F.S., Shah, M.: Transformers in vision: a survey. arXiv preprint arXiv:2101.01169 (2021)
10. Kodama, F., Fultz, P.J., Wandtke, J.C.: Comparing thin-section and thick-section CT of pericardial sinuses and recesses. Am. J. Roentgenol. **181**(4), 1101–1108 (2003)

11. Liang, J., Cao, J., Sun, G., Zhang, K., Van Gool, L., Timofte, R.: SwinIR: image restoration using swin transformer. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 1833–1844 (2021)
12. Liu, G., et al.: Recycling diagnostic MRI for empowering brain morphometric research-critical & practical assessment on learning-based image super-resolution. Neuroimage **245**, 118687 (2021)
13. Liu, Q., Zhou, Z., Liu, F., Fang, X., Yu, Y., Wang, Y.: Multi-stream progressive up-sampling network for dense CT image reconstruction. In: Martel, A.L., et al. (eds.) MICCAI 2020. LNCS, vol. 12266, pp. 518–528. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-59725-2_50
14. Liu, Z., et al.: Swin transformer: hierarchical vision transformer using shifted windows. arXiv preprint arXiv:2103.14030 (2021)
15. Lu, Z., Li, Z., Wang, J., Shi, J., Shen, D.: Two-stage self-supervised cycle-consistency network for reconstruction of thin-slice MR images. In: de Bruijne, M., et al. (eds.) MICCAI 2021. LNCS, vol. 12906, pp. 3–12. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-87231-1_1
16. Park, S., et al.: Computer-aided detection of subsolid nodules at chest CT: improved performance with deep learning-based CT section thickness reduction. Radiology **299**(1), 211–219 (2021)
17. Peng, C., Lin, W.A., Liao, H., Chellappa, R., Zhou, S.K.: Saint: spatially aware interpolation network for medical slice synthesis. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 7750–7759 (2020)
18. Peng, C., Zhou, S.K., Chellappa, R.: DA-VSR: domain adaptable volumetric super-resolution for medical images. In: de Bruijne, M., et al. (eds.) MICCAI 2021. LNCS, vol. 12906, pp. 75–85. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-87231-1_8
19. Shi, W., et al.: Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1874–1883 (2016)
20. Wang, Z., Chen, J., Hoi, S.C.: Deep learning for image super-resolution: a survey. IEEE Trans. Pattern Anal. Mach. Intell. **43**(10), 3365–3387 (2020)
21. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Process. **13**(4), 600–612 (2004)
22. Xu, F., et al.: Diagnostic performance of diffusion-weighted imaging for differentiating malignant from benign intraductal papillary mucinous neoplasms of the pancreas: a systematic review and meta-analysis. Front. Oncol. **11**, 2550 (2021)
23. Xuan, K., et al.: Reducing magnetic resonance image spacing by learning without ground-truth. Pattern Recognit. **120**, 108103 (2021)
24. Yang, J., et al.: *AlignShift*: bridging the gap of imaging thickness in 3D anisotropic volumes. In: Martel, A.L., et al. (eds.) MICCAI 2020. LNCS, vol. 12264, pp. 562–572. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-59719-1_55
25. Zhao, C., Dewey, B.E., Pham, D.L., Calabresi, P.A., Reich, D.S., Prince, J.L.: Smore: a self-supervised anti-aliasing and super-resolution algorithm for MRI using deep learning. IEEE Trans. Med. Imaging **40**(3), 805–817 (2020)
26. Zhou, S.K., et al.: A review of deep learning in medical imaging: imaging traits, technology trends, case studies with progress highlights, and future promises. Proc. IEEE **109**(5), 820–838 (2021)
27. Zhou, S., Zhang, J., Zuo, W., Loy, C.C.: Cross-scale internal graph neural network for image super-resolution. arXiv preprint arXiv:2006.16673 (2020)