Domain-Distance Adapted Super-Resolution Reconstruction of 1.5T MR Brain Images

Shan Cong[1,2], Yuzun Yang[1], Yang Zhou[3], Xinxin Wang[3], Haoran Luo[2], Hong Liang[2], and Xiaohui Yao[1,2]

[1] Qingdao Innovation and Development Center, Harbin Engineering University, Qingdao 266000, China
[2] College of Intelligent Science Systems and Engineering, Harbin Engineering University, Harbin 150001, China
{Shan.Cong, Xiaohui.Yao@hrbeu.edu.cn}
http://cnail.hrbeu.edu.cn/
[3] Harbin Medical University Cancer Hospital, Harbin 150000, China

Abstract. Fast magnetic resonance imaging (MRI) sequences are highly demanded in clinical settings. However, the inadequate imaging information can cause diagnostic difficulties. MR image super-resolution (SR) is a promising way to address this issue, but its performance is limited due to the practical difficulty of acquiring paired low- and high-resolution (LR and HR) images. Most existing methods use downsampled LR images, which can be inaccurate due to overlooking domain distance or poor approximation caused by unknown and complex degradation. In this study, we propose a domain-distance adapted SR framework for 1.5T MR brain images based on real but unpaired HR/LR images. Our framework leverages the ability to learn abstract representations of arbitrary unpaired images and adapt to the domain gap, making it feasible to identify realistic down-sampling. Additionally, we present a novel generative adversarial network (GAN) model that integrates a generator comprising an encoder, backbone, and decoder, along with an attention UNet-based discriminator and multi-scale perceptual loss. This approach yields visually convincing textures and successfully restores outdated 1.5T MRI data on a well-known public dataset, outperforming state-of-the-art SR approaches in both perceptual and quantitative evaluations.

Keywords: Domain-distance adapted down-sampling · MRI Super Resolution Reconstruction · Generative adversarial network

1 Introduction

Single image super-resolution (SR) enhances a low-resolution (LR) image by improving signal-to-noise ratio, accentuating texture details, and eliminating visual artifacts. These promising features are highly desirable in the field of brain imaging because 1) acquiring high-quality MR images can be expensive, uncomfortable for the patient, and motion-sensitive due to high field strengths and

extended scanning durations; and 2) analyzing substandard MR images with inadequate information is labor-intensive, demanding, and requires professional expertise, restricting diagnostic accuracy. Recent advances in deep learning-based MR image SR have shown promising results in improving patient experience and scanner efficiency while maintaining diagnostic quality [21]. However, several challenges must be addressed before clinical implementation.

One major challenge is modeling the unknown and complex degradation of a real LR image in the absence of paired LR/HR MR images. Since paired LR/HR MR images are usually unavailable, existing methods are dedicated to generating an LR/HR pair by synthesizing an LR image from the HR image to model the transformation [26,29]. Recent works primarily focused on explicit modeling such as bicubic down-sampling [9], k-space filtering [31], and a combination of image degradation operations [12,26]. These works performed well when the true degradation is known in prior. However, the performance is significantly compromised when the real degradation differs from their estimation. Alternatively, a few other works employ implicit modeling, such as domain-distance adaptation using convolutional neural network (CNN) [29], data distribution learning with GAN [7], and representation learning [25]. Compared to explicit modeling, implicit methods offer greater flexibility in handling various degradation scenarios, which motivates us to employ implicit strategy in this study.

Another challenge is recovering texture details while preserving the topological structures. An enhanced image may pass quantitative assessment but fail visual inspection as existing algorithms usually fail in balancing sharpness and structural details [33]. In some cases, tissue structure is even altered, which can be a fatal error as the structural texture is critical for visual diagnostics (i.e. lesion localization) and neuroscience studies (i.e. brain parcellation). There are two common directions to prioritize the preservation of topological structures and texture details: 1) design proper deep learning architectures for SR, where recent efforts have been made to improve autoregressive models [24], GAN [27,1], variational autoencoder (VAE) [13,3], encoding-decoding networks [2,25], and diffusion probabilistic models [10]; and 2) design loss functions to consider visual effects and yield convincing results together, for example, perceptual loss [27,9] and texture matching loss [5,22] functions. Overall, the investigation in both directions offers exciting opportunities for advancing MR image SR, and we believe that their combination has the potential to yield even more impressive results.

To tackle the aforementioned challenges, we aim to develop a realistic downsampling strategy and an effective texture restoration framework for brain MR image SR. Our contributions can be summarized as follows:

- We propose a transformer-based domain-distance adaptation strategy to bridge the domain gap between the downsampled and real LR images.
- We construct a GAN with an attention mechanism and perceptual loss, which leverages multi-scale image features to produce more biologically reasonable textures and precise edge contours.

– Our method enhances outdated 1.5T brain MR images and outperforms state-of-the-art SR approaches in both visual and quantitative evaluations, enabling the reuse of outdated datasets.

2 Methodology

We propose a two-stage framework to enhance MR image quality while preserving brain tissue structure and texture details. The down-sampling stage (Fig. 1(a, c)) employs a transformer architecture with content-aware positional encoding to capture invariant brain structural representation from HR images and avoid missing fine-grained details. The SR stage (Fig. 1(b)) uses a GAN architecture, where the generator (Fig. 1(d)) employs an encoder-decoder scheme and the discriminator (Fig. 1(e)) adopts an attention U-Net. Meanwhile, to ensure visually convincing textures, we incorporate multi-level perceptual loss into the SR stage. We present the domain-distance adapted down-sampling strategy in Sect. 2.1, the GAN-based SR framework in Sect. 2.2, 2.3 and 2.4, and the implementation details in Sect. 2.5 for reproducibility.

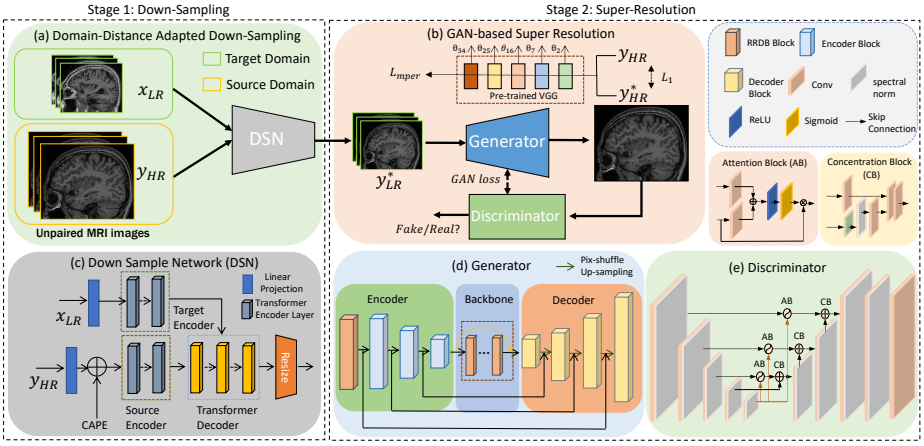

Fig. 1. Overview of the proposed two-stage framework. Stage 1 (a) involves a transformer-based down-sampling network (DSN), while Stage 2 (b) utilizes a GAN-based super resolution network (SRN).

2.1 Domain-Distance Adapted Down-Sampling

As shown in Fig. 1(c), we employ the transformer blocks with self-attention mechanism to extract image domain characteristics, textures, and structural features, facilitating the construction of source and target feature embeddings in the latent space. This mechanism enables the transformer to capture local and

global contextual information and generate multi-scale feature representation. In addition, we apply the content-aware positional encoding (CAPE) [4] on the high resolution image (y_{HR}) during multi-scale down-sampling, which aims to ensure that the brain structure and texture information remains invariant during the transformation process by encoding the positional information.

CAPE. Given a source image $y_{HR} \in R^{H_s \times W_s}$ and a target image $x_{LR} \in R^{H_t \times W_t}$, patches are projected into a sequential feature embedding $\mathcal{E}$ of size $L \times C$ using a linear projection layer, where $L = \frac{H}{m} \times \frac{W}{m}$, C is the dimension of $\mathcal{E}$, and the token is a $m \times m$ image patch. The CAPE P_{CA} between patches (i, j) is computed as follows:

$$\mathcal{P}_{\mathcal{L}} = \mathcal{F}_{\text{pos}}\left(\text{AvgPool}_{n \times n}(\mathcal{E})\right), \tag{1}$$

$$\mathcal{P}_{\mathcal{CA}}(i, j) = \sum_{k=0}^{s} \sum_{l=0}^{s} \left(a_{kl} \mathcal{P}_{\mathcal{L}}\left(i_k, j_l\right)\right), \tag{2}$$

where CAPE includes a 1×1 convolution operation $Fpos$, a learnable relative positional relationship P_L, a block size n, an interpolation weight a_{kl}, and the number of adjacent patches s.

Transformer Encoder. The transformer encoder consists of two encoders that encode information from both the target domain image x_{LR} and the source domain image y_{HR}. Each encoder is composed of a multi-head self-attention (MSA) module and a feed-forward network (FFN) [19]. The encoded feature ξ is obtained as follows:

$$\begin{aligned} \xi' &= F_{MSA}(Q, K, V) + Q, \\ \xi &= F_{FFN}(\xi') + \xi', \end{aligned} \tag{3}$$

Transformer Decoder. As illustrated in Fig. 1 (c), we take two encoded sequences ξ_x and ξ_y as input for the transformer decoder, which consists of two MSA layers and one FFN. Then the queries, keys, and values are calculated as $Q = \widehat{\xi_y} W_q,\ K = \xi_x W_k,\ V = \xi_x W_v$. The output $\hat{y}_{LR}$ can be calculated as:

$$\begin{aligned} \hat{y}''_{LR} &= F_{MSA}(Q, K, V) + Q, \\ \hat{y}'_{LR} &= F_{MSA}(\hat{y}''_{LR} + P_{CA}, K, V) + \hat{y}''_{LR} \\ \hat{y}_{LR} &= F_{FFN}\left(\hat{y}'_{LR}\right) + \hat{y}'_{LR} \end{aligned} \tag{4}$$

Resizing. For the training purposes, the decoded output $\hat{y}_{LR}$ undergoes additional processing through a three-layer up-sampling decoder [32] and a linear degradation operation to obtain the downsampled result y^*_{LR}.

Objective Function. The loss function for the source domain (L_s) preserves the original tissue structure and texture of the MRI, ensuring fidelity of the generated results. The loss function for the target domain (L_t) incorporates image characteristics captured from the target domain, facilitating domain adaptation for high-quality outputs that are faithful to the low-resolution input. The total loss function of DSN is calculated as:

$$\mathcal{L} = \lambda_t \mathcal{L}_t + \lambda_s \mathcal{L}_s + \lambda_{id1} \mathcal{L}_{id1} + \lambda_{id2} \mathcal{L}_{id2}, \tag{5}$$

where L_s is calculated as:

$$\mathcal{L}_s = \frac{1}{N_l} \sum_{i=0}^{N_l} \left\| \phi_i \left(y_{LR}^* \right) - \phi_i \left(y_{HR} \right) \right\|_2 , \tag{6}$$

and L_t is calculated as:

$$\begin{aligned} \mathcal{L}_t = \frac{1}{N_l} \sum_{i=0}^{N_l} & \left\| \mu \left(\phi_i \left(y_{LR}^* \right) \right) - \mu \left(\phi_i \left(x_{LR} \right) \right) \right\|_2 \\ & + \left\| \sigma \left(\phi_i \left(y_{LR}^* \right) \right) - \sigma \left(\phi_i \left(x_{LR} \right) \right) \right\|_2 , \end{aligned} \tag{7}$$

where N_i is the number of layers, $\mathcal{L}_{id1}$ and $\mathcal{L}_{id2}$ are calculated following the training strategy described in [18], $\phi_i(.)$ is the feature of the ith layer extracted from pre-trained VGG networks (described in the Sect. 2.5).

2.2 SR Network: Generator Architecture

Encoder. We employ the residual-in-residual dense block (RRDB) [27] for effective feature extraction, which first extracts features from the low-resolution MR image, yielding a low-dimensional feature representation f_0. Then, in order to capture multi-scale feature information f_i, we subsequently employ convolutional modules to gradually decrease the resolution of the features while increasing their dimensionality. Thus, we obtain $f_i = E_i \left(f_{i-1} \right), i \in \{1, \cdots, N\}$, where $E_i(\cdot)$ is the convolutional layer with the kernel size of 4, stride of 2, and padding of 1.
Backbone. We leverage previous findings [11,30,23] that suggest the incorporation of dense connections and deeper networks. To achieve deep feature fusion and extraction, we use cascaded RRDB blocks in high-dimensional space. This method allows for the fusion of high-dimensional features before and after the network is deepened through dense connections, meanwhile using residual blocks to alleviate computational overhead and gradient vanishing issues from deeper networks.
Decoder. A decoder with progressive fusion is used to merge encoder and backbone features for HR MR image generation. The decoder uses convolutional layers for upsampling and pixshuffle for sub-pixel convolution. Skip connections are used to retain low-dimensional information and perform multi-scale fusion, which allows the model to simultaneously focus on both the large-scale contour structures and small-scale texture details, resulting in more detailed and accurate MR image reconstruction.

2.3 SR Network: Discriminator Architecture

Inspired by [17], we adopt an attention U-net discriminator architecture in this work to focus on poorly generated texture during adversarial training. The architecture includes concentration and attention blocks, and a U-net structure, as shown in Fig.1 (e). To improve training stability and make the discriminator

more suitable for image SR tasks, we use a spectral normalization regularization operation described in [16]. The loss function is calculated as:

$$L_D = \sum_{i=1}^{W}\sum_{j=1}^{H}\left(-E\left[\log\left(D(y_{HR}, y_{HR}^{*})\left[i,j\right]\right)\right]\right. \\ \left.-E\left[1-\log\left(D\left(y_{HR}^{*}, y_{HR}\right)\left[i,j\right]\right)\right]\right), \tag{8}$$

where W and H are the width and height of the image input to the discriminator.

2.4 SR Network: Objective Function

To preserve the structural information of MRI and enhance its texture details, we use generator loss which includes multi-perceptual loss L_{mper}, 1-norm distance loss L_1, and GAN loss L_{gan} [6]. The total loss function is calculated as:

$$L_{SRN} = \alpha L_{mper} + \lambda L_1 + \eta L_{gan}, \tag{9}$$

where α, λ and η are weights, L_{mper} is shown in the Fig. 1 (b) and detailed settings are described in Sect. 2.5.

2.5 Details of Implementation

Our proposed model is trained on 2 NVIDIA GeForce RTX 3090 GPUs with 48 GB memory using the PyTorch framework. The downsample network is trained for 300,000 iterations with a batch size of 3 and hidden dimension of 512. For the SRN, a cosine annealing with warm restarts learning rate optimization strategy is employed with a minimum learning rate of 1e-7, and the total number of iterations is set to 250,000. VGG-19 is used to extract image features for the multi-layer perceptual loss L_{mper}, with layers 2, 7, 16, 25, and 34 selected and weighted at 0.1, 0.1, 1.0, 1.0, and 1.0, respectively. The weights for the L_1 loss and perceptual loss are both set to 1, while the weight for the GAN loss is set to 0.1.

3 Examples and Applications

3.1 Data Collection And Evaluation Protocol

ADNI LR data. A total of 41 T1-weighted 1.5T MRI scans from the publicly available Alzheimer's Disease Neuroimaging Initiative Phase-1 (ADNI-1) database are used in this study. The original resolution is $256 \times 256 \times 160$, with 22 subjects used for training, and 19 for testing. It should be noted that the ADNI-1 dataset is generally considered to have lower image quality compared to the later phases of the ADNI study. Detailed data acquisition protocols can be accessed at https: //adni.loni.usc.edu.

Private HR/LR paired data. Three volunteers undergo both HR and LR MRI scans on a 3.0T GE scanner at our site using a 3D T1-weighted MP-RAGE

sequence. The high-resolution scans have a slice thickness/gap of 1/0 mm, spatial resolution of 1.0×1.0×1.0 mm, TE of 3.444 ms, TR of 2737.62 ms, and a scanning time of 6 min 48s. The same sequence is used for the low-resolution scanning with a reduced TE of 1.66 ms and scanning time of 4 min 28s.

Evaluation metric. When ground truth is available (Sect.3.3), PSNR [20] and SSIM [28] are used to evaluate the down-sampled image quality. In the absence of reference images as ground truth (Sect.3.2), NIQE [15] and BRISQUE [14] are applied to compare the enhanced image quality. Metrics are reported as averaged values in the following sections.

3.2 Super-Resolution on ADNI-1 data

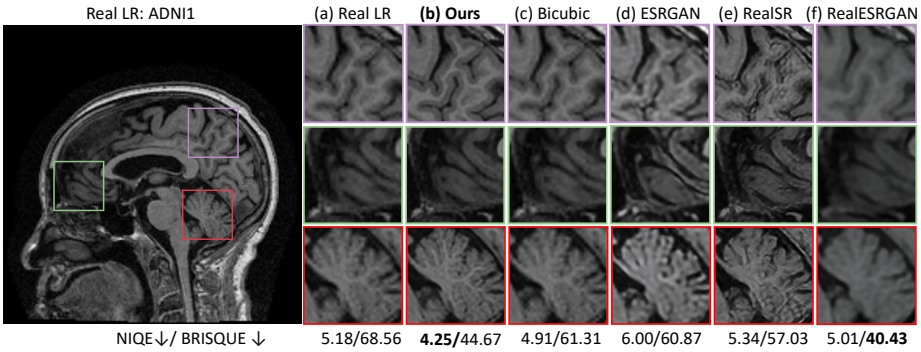

Fig. 2. Comparison of the image reconstruction methods. A smaller value of NIQE/BRISQUE index indicates higher image quality.

We apply the proposed framework on the 1.5T MRI data from ADNI-1 (Sect.3.1), acknowledged to have lower quality. To assess the effectiveness of our approach, we use non-reference metrics (NIQE and BRISQUE), and conduct a comparative analysis with benchmark method Bicubic and state-of-the-art methods including ESRGAN[27], RealSR [8], and RealESRGAN [26]. As illustrated in Fig. 2, our method outperforms others in terms of the NIQE metric and achieves the second-best result in the BRISQUE index.

We further analyze the reasons for visual discrepancies among SR methods on the ADNI-1 dataset. Bicubic upsampling produces the worst BRISQUE, potentially because it simply interpolates the image without improving its quality, while ESRGAN shows the worst NIQE potentially due to domain gap. RealSR achieves good visual results in certain regions, but with structural alterations and unnecessary texture in other regions. RealESRGAN achieves the best BRISQUE, but sacrifices texture details and suffers from the artifacts, such as the expansion of cortical surface thickness and the reduction of sulcal depth of the brain. The proposed method successfully enhances image quality by preserving the structure and recovering biologically significant texture information simultaneously.

3.3 Evaluation on Down-Sampling

We compare the proposed domain distance-adaptive down-sampling (DDADS) method with down-sampling methods adopted in the RealSR and RealESRGAN, and K-space method for domain adaptation using three LR/HR image pairs from the private dataset since no paired LR and HR images are available from ADNI. Evaluation metrics include PSNR and SSIM to quantitatively measure the difference between the synthesized image and the ground truth.

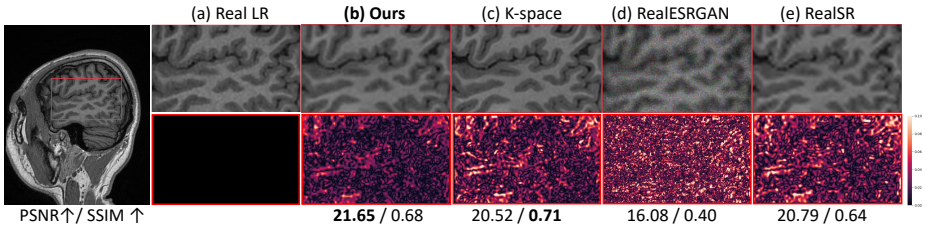

Fig. 3. Comparison of the down-sampling methods. A higher PSNR/SSIM index value indicates higher image quality, while a lower heat-map value indicates a smaller distance between the synthesized image and the real LR.

As depicted in Fig.3, the proposed DDADS method outperforms RealSR, K-space, and RealESRGAN in terms of the heatmap and achieves the highest PSNR score, as well as the second-best SSIM score. K-space has the highest SSIM score but generates some artifact texture details. RealSR performs slightly worse than our method across all evaluation metrics. RealESRGAN produces the worst PSNR and SSIM scores, possibly due to its down-sampling strategy being designed for natural images.

It is worth noting that to ensure fairness in comparison experiments, all models are retrained on our datasets using the same and best-performed model architecture and hyperparameters as described in their respective papers. Specifically, ESRGAN uses the K-space method for down-sampling, while RealSR uses a degradation framework to estimate blur kernels, and RealESRGAN utilizes the high-order degradation modeling method.

4 Conclusion

This paper proposes a novel domain-distance adapted super-resolution framework for 1.5T MR brain images. Our approach combines a transformer-based domain adaptation network, an encoding-decoding generator architecture, an attention UNet-based discriminator for SR, and perceptual loss to effectively preserve brain structural and texture information. Our experimental results demonstrate the superiority of our proposed model compared to state-of-the-art SR approaches in both quantitative evaluation metrics and perceptual quality

assessments. This indicates that our proposed framework is a reliable solution for restoring outdated MRI images and has the potential to enhance clinical MR images. The proposed approach has wide applicability to various medical imaging tasks and can assist medical professionals in obtaining more accurate and detailed diagnostic information from 1.5T MR brain images. Therefore, our work has noteworthy contributions to the field of medical imaging and could have a positive impact on patient care.

Acknowledgement. ADNI data collection and sharing for this project is funded by the Alzheimer's Disease Metabolomics Consortium (National Institute on Aging R01AG046171, RF1AG051550 and 3U01AG024904-09S4). For private HR data, informed consent has been obtained from all participants and the research ethics committees or institutional review boards approved the data collection.

References

1. Bell-Kligler, S., Shocher, A., Irani, M.: Blind super-resolution kernel estimation using an internal-gan. Advances in Neural Information Processing Systems **32** (2019)
2. Chan, K.C., Wang, X., Xu, X., Gu, J., Loy, C.C.: Glean: Generative latent bank for large-factor image super-resolution. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 14245–14254 (2021)
3. Chira, D., Haralampiev, I., Winther, O., Dittadi, A., Liévin, V.: Image super-resolution with deep variational autoencoders. In: Computer Vision–ECCV 2022 Workshops: Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part II. pp. 395–411. Springer (2023)
4. Deng, Y., Tang, F., Dong, W., Ma, C., Pan, X., Wang, L., Xu, C.: Stytr2: Image style transfer with transformers. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11326–11336 (2022)
5. Gondal, M.W., Schölkopf, B., Hirsch, M.: The unreasonable effectiveness of texture transfer for single image super-resolution. In: Computer Vision–ECCV 2018 Workshops: Munich, Germany, September 8-14, 2018, Proceedings, Part V 15. pp. 80–97. Springer (2019)
6. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial networks. Communications of the ACM **63**(11), 139–144 (2020)
7. Huang, B., Xiao, H., Liu, W., Zhang, Y., Wu, H., Wang, W., Yang, Y., Yang, Y., Miller, G.W., Li, T., et al.: Mri super-resolution via realistic downsampling with adversarial learning. Physics in Medicine & Biology **66**(20), 205004 (2021)
8. Ji, X., Cao, Y., Tai, Y., Wang, C., Li, J., Huang, F.: Real-world super-resolution via kernel estimation and noise injection. In: proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops. pp. 466–467 (2020)
9. 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)
10. Li, H., Yang, Y., Chang, M., Chen, S., Feng, H., Xu, Z., Li, Q., Chen, Y.: Srdiff: Single image super-resolution with diffusion probabilistic models. Neurocomputing **479**, 47–59 (2022)

11. Lim, B., Son, S., Kim, H., Nah, S., Mu Lee, K.: Enhanced deep residual networks for single image super-resolution. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops. pp. 136–144 (2017)
12. Liu, C., Sun, D.: On bayesian adaptive video super resolution. IEEE transactions on pattern analysis and machine intelligence **36**(2), 346–360 (2013)
13. Liu, Z.S., Siu, W.C., Chan, Y.L.: Photo-realistic image super-resolution via variational autoencoders. IEEE Transactions on Circuits and Systems for video Technology **31**(4), 1351–1365 (2020)
14. Mittal, A., Moorthy, A.K., Bovik, A.C.: No-reference image quality assessment in the spatial domain. IEEE Trans on image processing **21**(12), 4695–4708 (2012)
15. Mittal, A., Soundararajan, R., Bovik, A.C.: Making a "completely blind" image quality analyzer. IEEE Signal processing letters **20**(3), 209–212 (2012)
16. Miyato, T., Kataoka, T., Koyama, M., Yoshida, Y.: Spectral normalization for generative adversarial networks. In: International Conference on Learning Representations (2018)
17. Oktay, O., Schlemper, J., Folgoc, L.L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N.Y., Kainz, B., et al.: Attention u-net: Learning where to look for the pancreas. arXiv preprint arXiv:1804.03999 (2018)
18. Park, D.Y., Lee, K.H.: Arbitrary style transfer with style-attentional networks. In: proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 5880–5888 (2019)
19. Parmar, N., Vaswani, A., Uszkoreit, J., Kaiser, L., Shazeer, N., Ku, A., Tran, D.: Image transformer. In: International conference on machine learning. pp. 4055–4064. PMLR (2018)
20. Poobathy, D., Chezian, R.M.: Edge detection operators: Peak signal to noise ratio based comparison. IJ Image, Graphics and Signal Processing **10**, 55–61 (2014)
21. Rudie, J.D., Gleason, T., Barkovich, M.J., Wilson, D.M., Shankaranarayanan, A., Zhang, T., Wang, L., Gong, E., Zaharchuk, G., Villanueva-Meyer, J.E.: Clinical assessment of deep learning–based super-resolution for 3d volumetric brain mri. Radiology: Artificial Intelligence **4**(2) (2022)
22. Sajjadi, M.S., Scholkopf, B., Hirsch, M.: Enhancenet: Single image super-resolution through automated texture synthesis. In: Proceedings of the IEEE international conference on computer vision. pp. 4491–4500 (2017)
23. Tong, T., Li, G., Liu, X., Gao, Q.: Image super-resolution using dense skip connections. In: Proceedings of the IEEE international conference on computer vision. pp. 4799–4807 (2017)
24. Van Den Oord, A., Kalchbrenner, N., Kavukcuoglu, K.: Pixel recurrent neural networks. In: Inter. conf. on machine learning. pp. 1747–1756. PMLR (2016)
25. Wang, L., Wang, Y., Dong, X., Xu, Q., Yang, J., An, W., Guo, Y.: Unsupervised degradation representation learning for blind super-resolution. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 10581–10590 (2021)
26. 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)
27. 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 ECCV workshops. pp. 0–0 (2018)
28. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. IEEE transactions on image processing **13**(4), 600–612 (2004)

29. Wei, Y., Gu, S., Li, Y., Timofte, R., Jin, L., Song, H.: Unsupervised real-world image super resolution via domain-distance aware training. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 13385–13394 (2021)
30. Zhang, Y., Li, K., Li, K., Wang, L., Zhong, B., Fu, Y.: Image super-resolution using very deep residual channel attention networks. In: Proceedings of the European conference on computer vision (ECCV). pp. 286–301 (2018)
31. 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 transactions on medical imaging **40**(3), 805–817 (2020)
32. Zheng, S., Lu, J., Zhao, H., Zhu, X., Luo, Z., Wang, Y., Fu, Y., Feng, J., Xiang, T., Torr, P.H., et al.: Rethinking semantic segmentation from a sequence-to-sequence perspective with transformers. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 6881–6890 (2021)
33. Zhou, W., Jiang, Q., Wang, Y., Chen, Z., Li, W.: Blind quality assessment for image superresolution using deep two-stream convolutional networks. Information Sciences **528**, 205–218 (2020)