Probabilistic Prior-Guided Anatomical Alignment for MRI Super-Resolution

Yiwen Luo[1], Xiaoying Tang[2], and Yixuan Yuan[3(✉)]

Department of Electrical Engineering, City University of Hong Kong
Department of Electrical and Electronic Engineering, Southern University of Science and Technology
Department of Electronic Engineering, The Chinese University of Hong Kong
yxyuan@ee.cuhk.edu.hk

Abstract. High-resolution (HR) magnetic resonance imaging (MRI) offers exceptional visualization of human tissue but is often limited by hardware constraints. While recent super-resolution (SR) methods leveraging learned codebooks have shown promise, they often overlook the rich anatomical priors inherent in MRI data. To address this, we propose a probabilistic prior-guided anatomical alignment for MRI super-resolution (PGASR) method that incorporates anatomical knowledge into the SR process. Specifically, we first introduce an anatomical-conditioned codebook generation (ACG) module that generates rough anatomical structure maps by extracting the regions of interest from MRI slices. These maps are used as anatomical conditions for the discrete codebook generation. Then, to better exploit information between MRI slices, we propose a prior matching alignment (PMA) module that aligns the codebook index matching probabilities between adjacent slices, as well as across low-resolution (LR) and high-resolution (HR) domains, thereby reducing the loss of image details. We validate the effectiveness of the proposed PGASR method with the public MRI dataset IXI. The experimental results demonstrate that PGASR outperforms state-of-the-art methods.

Keywords: Super Resolution · Magnetic Resonance Imaging · Prior-Guided Codebook Alignment.

1 Introduction

Magnetic resonance imaging (MRI) [1–3] is widely utilized in clinical practice for diagnosis and image-guided therapy due to its superior soft tissue contrast and non-invasive nature. However, achieving high-resolution MRI is often constrained by acquisition time, magnetic field strength, and high costs. These limitations compromise the texture and structural details essential for accurate diagnosis and quantitative analysis. Image super-resolution (SR) [4] has emerged as a promising solution for addressing this problem, which aims to restore high-resolution (HR) images from their degraded low-resolution (LR) observations. Image super-resolution methods can be broadly categorized into: reconstruction-

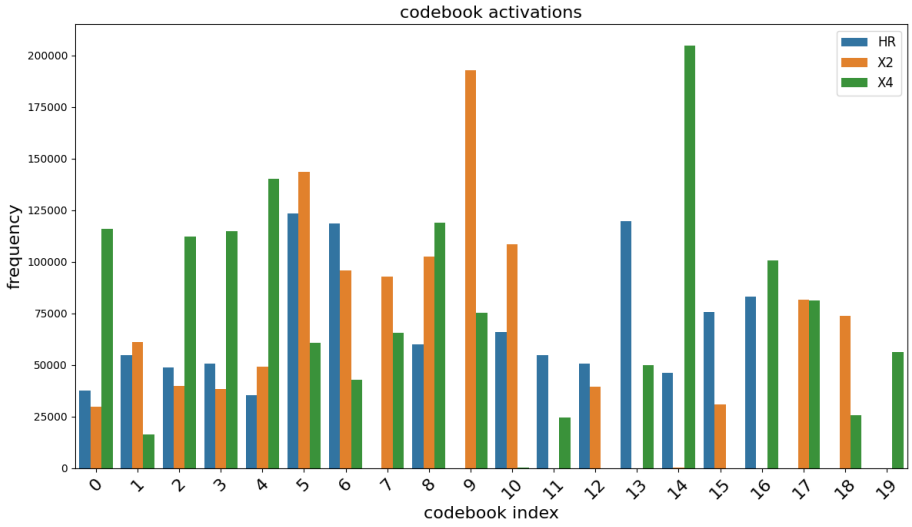

Fig. 1. The codebook activations in brain IXI dataset.

based methods [5], which solve ill-posed problems via regularization; learning-based methods [6], leveraging deep learning models to learn mappings from low to high-resolution images; and example-based methods [7], which utilize external data to establish mappings of HR and LR images.

Recent advancements in codebook-based SR techniques [6, 8, 9] have introduced innovative ways to model high-resolution textures by leveraging discrete feature spaces derived from pre-trained codebooks. These methods [6, 8, 9] constrain the feature output to a latent discrete space to ensure the generated textures are consistent with realistic image representations. However, current approaches typically overlook the rich anatomical information inherent in MRI data, which could serve as valuable priors for SR. MRI anatomical structures [10] exhibit distinct contrasts and textures due to variations in tissue composition, water content, and magnetic properties. Rough tissue segmentation in MRI can provide crucial anatomical insights for image reconstruction. This motivates the use of foundational segmentation models (*e.g.*, segment anything model [11]) to generate segments as rough anatomical structure maps for MRI slices. By conditioning the pretraining of the discrete latent space on anatomical priors, we develop an anatomical-conditioned codebook that enhances SR by improving texture and structure alignment with anatomical structures.

Moreover, existing methods [6, 8] typically quantize image features into discrete latent representations independently, failing to capture the inherent inter-slice correlations in volumetric MRI data. This lack of integration compromises structural continuity, which is crucial for accurate MRI reconstruction. Additionally, degraded LR images often suffer from loss of detail textures, leading to ambiguous feature distributions and inconsistent codebook mappings with HR

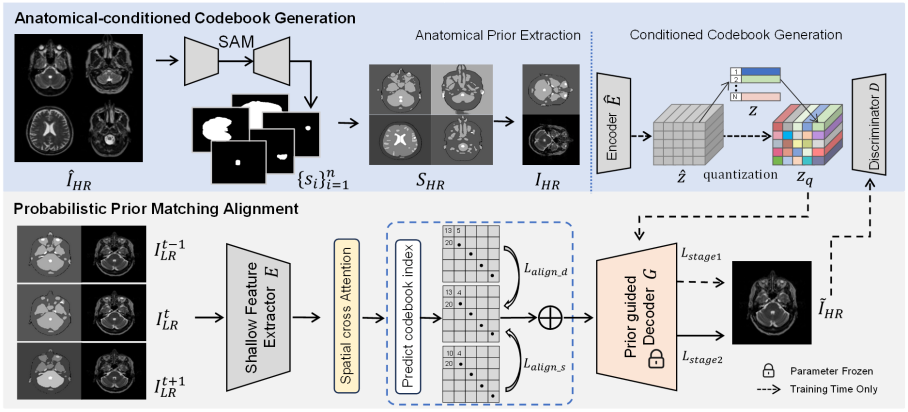

Fig. 2. The framework of the proposed PGASR.

images. As shown in Fig. 1, the VQGAN [6] trained on HR images from IXI dataset exhibits inconsistent codebook activation probabilities between HR images and their $\times 2/\times 4$ LR counterparts, especially in the code of 10 to 19, where corresponding activations are conspicuously absent. To minimize distortion in the reconstructed images, the codebook index matching probabilities during LR image quantization should ideally align closely with those of the HR images.

In this paper, we propose a probabilistic prior-guided anatomical alignment for MRI super-resolution (PGASR) to address the aforementioned limitations in SR. Specifically, we design an anatomical-conditioned codebook generation (ACG) module, which utilizes the foundational segmentation models to produce rough anatomical structure maps. These maps are then used to guide the generation of an anatomical-conditioned codebook, enabling the SR process to incorporate anatomical knowledge effectively. Then, we propose a prior matching alignment (PMA) module to enhance inter-slice feature consistency by aligning information across adjacent slices. We further constrain the matching probabilistic consistency in codebooks between HR and LR images, ensuring better feature alignment for MRI SR.

2 Method

The framework of our proposed *Probabilistic Prior-Guided Anatomical Alignment for MRI Super-Resolution (PGASR)* is illustrated in Fig. 2. In the first stage, we utilize a pre-trained SAM to extract anatomical structure maps by segmenting regions of interest in MRI. Given a HR image $\hat{I}_{HR}$, we generate segments $\{s_i\}_{i=0}^{n}$ to derive its anatomical prior S_{HR}. S_{HR} is then used as conditions to learn the discrete codebook Z, encoder $\hat{E}$, decoder G, and discriminator D. In the second stage, LR slices are first input into a randomly initialized shallow feature extractor E. The generated features are then fed into a spatial cross-attention module, which is optimized with the alignment losses $\mathcal{L}_{align\_d}$ and

$\mathcal{L}_{align\_s}$ to ensure codebook index matching consistency between adjacent slices and across the LR and HR domains. Finally, the quantized feature is reconstructed by the fixed decoder G. During inference, we employ the same modules as in the second stage, with all parameters fixed.

2.1 Anatomical-conditioned Codebook Generation

Anatomical Prior Extraction. Manual annotation of anatomical structures in MRI slices is labor-intensive and clinically impractical. To address this challenge, we employ the Segment Anything Model (SAM) [11] to extract anatomical priors in an unsupervised manner. Given a high-resolution MRI slice $\hat{I}_{HR}$, we leverage the pre-trained SAM to generate a set of spatially coherent segments $\{s_i\}_{i=1}^{n}$ that delineate critical brain regions, such as gray matter, white matter, and cerebrospinal fluid. These segments are then aggregated via a summation operation across the spatial dimension, followed by normalization to [0, 1]. The resulting composite segmentation map, denoted as S_{HR}, serves as a spatially grounded anatomical prior (Fig. 2, top-left). This automated process bypasses manual annotation while retaining structural fidelity for downstream tasks.

Conditioned Codebook Generation. To learn a discrete codebook conditioned on extracted anatomical priors, we employ a vqgan-like architecture with an encoder $\hat{E}$, a learnable codebook Z, and a decoder G. For a given HR image $\hat{I}_{HR}$, we concatenate it with its extracted anatomical prior S_{HR} to form the composite input I_{HR}, which is then fed into encoder $\hat{E}$ to derive feature embeddings $\hat{z} = E(I_{HR}) \in \mathbb{R}^{H\times W\times n_z}$. We predict the codebook index k for each entry $\hat{z}_{ij} \in \mathbb{R}^{n_z}$ in $\hat{z}$ by retrieving the nearest entries in codebook $Z \in \mathbb{R}^{N\times n_z}$, where N is codebook size and n_z is the dimension of codes. The quantized discrete representation of input $\hat{z}$ is formulated as,

$$z_{\mathbf{q}} = \left(\underset{z_k \in \mathcal{Z}}{\arg\min} \|\hat{z}_{ij} - z_k\| \right) \in \mathbb{R}^{H\times W\times n_z} \tag{1}$$

where z_k is the selected codebook entry of $\hat{z}_{ij}$. The quantized features $z_{\mathbf{q}}$ are finally decoded by the decoder G to reconstruct the HR image as

$$\tilde{I}_{HR} = G(z_q) \tag{2}$$

Unlike the original VQGAN encoder structure, we modify the architecture of $\hat{E}$ to a resnet-transformer architecture to reduce learning complexity. Specifically, we first extract the visual features using resnet blocks and flatten them to serve as input for the transformer block. To better capture the spatial and channel relationships within the MRI slice, we replace the standard transformer block with the DAT block [12]. The encoder E, codebook Z, and decoder G are adversarially trained with a discriminator D. In the first stage, we optimize the model using L1 loss, perceptual loss $\mathcal{L}_{\text{per}}$ and adversarial loss $\mathcal{L}_{\text{adv}}$. The total loss function is defined as,

$$\mathcal{L}_{stage1} = \mathcal{L}_1 + \mathcal{L}_{\text{per}} + \mathcal{L}_{\text{adv}} + \|\text{sg}[E(x)] - z_{\mathbf{q}}\|_2^2 + \beta \|\text{sg}[z_{\mathbf{q}}] - E(x)\|_2^2 \tag{3}$$

where $sg[\cdot]$ denotes the stop-gradient operation, and the $\|\mathrm{sg}\,[z_{\mathbf{q}}] - E(x)\|_2^2$ represents the commitment loss with a weighting factor β.

2.2 Prior Matching Alignment

To address the limitations of common codebook-based methods that independently match input visual features with learned codebook entries, we propose two constraints. First, we enforce inter-slice consistency in codebook index matching by constraining matching across adjacent slices. Second, to address the discrepancy in codebook index matching probabilistic between HR and LR domains (as evidenced in Fig. 1), we enforce consistency in the matching process across HR and LR slices.

Given an input LR MRI slice I_{LR}^t (which includes the original LR image and its corresponding anatomical prior), we denote its high-resolution counterpart as I_{HR}^t. Its adjacent slices are denoted as I_{LR}^{t-1} and I_{LR}^{t+1}. We first feed these three slices into a shallow feature extractor E (identical to the network structure of $\hat{E}$ but trained independently), which produces feature embeddings $[\hat{z}^{t-1}, \hat{z}^t, \hat{z}^{t+1}] \in \mathbb{R}^{H\times W\times n_z}$. At this stage, the codebook Z remains fixed. To exploit the inherent texture similarity among adjacent MRI slices, we introduce a spatial cross-attention block (SCAB) to enhance the feature embeddings. As illustrated in Fig. 3, SCAB is designed based on the dual spatial transformer block (DSTB) [12], which performs cross-attention by using $\hat{z}^t$ as the query, and $\hat{z}^{t-1}$ and $\hat{z}^{t+1}$ as the key and value, respectively. The features refined by SCAB are then processed for quantization using the codebook Z. For clarity, we denote the refined features from SCAB as $[\hat{z}^{t-1}, \hat{z}^t, \hat{z}^{t+1}]$. Since adjacent slices exhibit similar textures, they are expected to have similar representations in the codebook space. Denote the distances between the refined features and all codebook entries as $[a^{t-1}, a^t, a^{t+1}]$, we regard the predicted probabilities for codebook indices as distributions over the learned codebook and design the spatial alignment loss as,

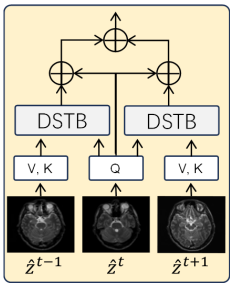

Fig. 3. Spatial cross attention block.

$$\mathcal{L}_{align\_s} = D_{KL}(a^{t-1}, a^t) + D_{KL}(a^t, a^{t+1}) \tag{4}$$

where $D_{KL}(\cdot,\cdot)$ is the KL divergencr. $\mathcal{L}_{align\_s}$ leverages spatial dependencies between adjacent slices to enforce consistency in codebook matching, improving feature alignment across adjacent MRI slices.

We further introduce $\mathcal{L}_{align\_d}$ to mitigate the discrepancy in the codebook index matching probabilistic between HR and LR domains. Specifically, we pass I_{HR}^t through the learned encoder $\hat{E}$ (see Sec. 2.1) and predict the codebook index probabilities as a_{HR}^t. The domain alignment loss $\mathcal{L}_{align\_d}$ is then denoted as $\mathcal{L}_{align\_d} = D_{KL}(a^t, a_{HR}^t)$. While $\mathcal{L}_{align\_s}$ improves inter-slice information alignment, $\mathcal{L}_{align\_d}$ ensures consistent codebook index matching between HR and LR domains. The overall loss for the second stage is given by,

$$\mathcal{L}_{stage2} = \mathcal{L}_1 + \alpha_1\mathcal{L}_{align\_s} + \alpha_2\mathcal{L}_{align\_d} \tag{5}$$

where α_1 and α_2 are hyperparameters. Note that the Decoder G and codebook Z remain fixed during this stage.

3 Experiments

3.1 Experimental setup

Dataset. We evaluate our method on T2-weighted MRI brain volumes from the IXI dataset, using 300 subjects for training and 176 subjects for testing. Each volume contains axial slices of size 256×256 pixels (1 mm isotropic resolution). To mitigate noise and intensity inhomogeneity in frontal and posterior regions, we exclude the first and last few slices of each volume, retaining 80 standardized slices per subject. HR slices are normalized to zero-mean and unit variance. To generate low-resolution (LR) counterparts, we degrade HR images by first applying a general anisotropic Gaussian blur kernel, followed by additive white Gaussian noise, and finally bicubically downsampling the degraded image at scale factors of $\times 2$ and $\times 4$.

Evaluation Metrics. We utilized the peak signal-to-noise ratio (PSNR), structural similarity index (SSIM), and learned perceptual image patch similarity (LPIPS) commonly used in super-resolution to evaluate the reconstructed image quality. The main experimental results are averaged over 3 runs.

Implementation Detail. Our proposed framework is implemented using the PyTorch library. In the first stage, we optimize the encoder $\hat{E}$, codebook Z, decoder G, and discriminator D with an initial learning rate of 0.0001, which is halved every 10 epochs. The batch size for this stage is set to 16. The hyperparameter β in $\mathcal{L}_{stage2}$ is set to 0.25 and the number of codes in codebooks is set to 256 for all experiments. In the second stage, we optimize the feature extractor E and spatial cross- attention block with an initial learning rate of 0.0003 and a batch size of 8. The loss weights α_1 and α_2 in $\mathcal{L}_{stage2}$ are empirically set to 0.1. Architecturally, both encoder $\hat{E}$ and E adopt a 4-level design: the first two blocks use ResNet for local feature extraction, while the latter two employ the DAT block for adaptive interaction in spatial and channel. The decoder G consists of four sequential DAT blocks. All the compared methods are retrained using their default parameter settings for fair comparison.

3.2 Experiment results

To validate the effectiveness of the proposed PGASR, we conduct experiments on IXI dataset with scale factors of $\times 2$ and $\times 4$. We compared our approach with state-of-the-art SR methods for common images [13, 14, 12], as well as with methods specifically designed for MRI SR [2, 3]. All methods follow the original experimental settings to ensure a fair comparison of performance. The comparison results are shown in Table 1.

Table 1. Quantitative comparison results for image SR on brain IXI dataset. Best and second best results are **highlighted** and underlined

method	×2			×4		
	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓
Bicubic	28.5198	0.7334	0.0892	25.9486	0.7268	0.1032
T2Net [2]	32.8744	0.9311	0.0749	28.6592	0.8549	0.0987
MINet [3]	33.7914	0.9426	0.0758	28.8441	0.8677	0.0946
ART [13]	36.2201	0.9518	0.0781	30.1816	0.8800	0.1155
CAT [14]	36.3455	0.9537	0.0682	<u>30.4928</u>	<u>0.8891</u>	<u>0.0970</u>
DAT [12]	<u>36.4619</u>	<u>0.9598</u>	<u>0.0642</u>	30.4429	0.8817	0.1078
Ours	**36.7642**	**0.9706**	**0.0487**	**30.6844**	**0.8972**	**0.0924**

Table 2. Ablation study. Best results are **highlighted**

Exp.	ACG	$\mathcal{L}_{align\_d}$	$\mathcal{L}_{align\_s}$	SCAB	×2			×4		
					PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS
0	-	-	-	-	34.487	0.9527	0.0613	28.955	0.8717	0.0953
1	✓	-	-	-	35.942	0.9608	0.0575	29.641	0.8794	0.0948
2	✓	✓	-	-	36.428	0.9661	0.0547	29.726	0.8869	0.0927
3	✓	-	✓	-	36.254	0.9648	0.0563	29.662	0.8834	0.0941
4	✓	✓	✓	-	36.597	0.9679	0.0504	30.437	0.8945	0.0931
Ours	✓	✓	✓	✓	**36.764**	**0.9706**	**0.0487**	**30.852**	**0.8972**	**0.0924**

The evaluation result on IXI. As demonstrated by the IXI experiment results in Table 1, our method achieves superior performance compared to existing techniques across both fidelity (PSNR/SSIM) and perceptual quality (LPIPS) metrics. For the ×2 super-resolution task, our approach attains state-of-the-art results with PSNR and SSIM values of 36.7642 dB and 0.9706, respectively, outperforming the previous best method (6^{th} row) by 0.3023 (PSNR) and 0.0108 (SSIM). Notably, it achieves the lowest LPIPS score of 0.0487, demonstrating a 26.7% improvement in perceptual quality over the closest competitor (DAT: 0.0642). For the ×4 task, our method maintains its superiority with 30.6844 dB in PSNR and 0.8972 in SSIM, outperforming the second-best approach by 0.1916 (PSNR) and 0.0081 (SSIM), while consistently achieving lower LPIPS scores. These results validate the capability of our method to balance fidelity and perceptual realism, highlighting its robustness across varying degradation levels. The consistent improvements underscore the practical utility of our framework for MRI SR applications.

Ablation study. We conduct an ablation study to further investigate the validity of each component in PGASR. The results are shown in Table. 2, "ACG" refers to the use of ACG in the first stage. "$\mathcal{L}_{align\_d}$" and "$\mathcal{L}_{align\_s}$" denote the incorporation of the alignment loss in $\mathcal{L}_{stage2}$. "SCAB" stands for the use of the Spatial Cross Attention Block. The results of the ablation study show that the

Table 3. KL divergence of codebook activations between HR and LR

Setting	x2	x4
w/o L_{align_d}	0.0597	0.0861
with L_{align_d}	0.0413	0.0530

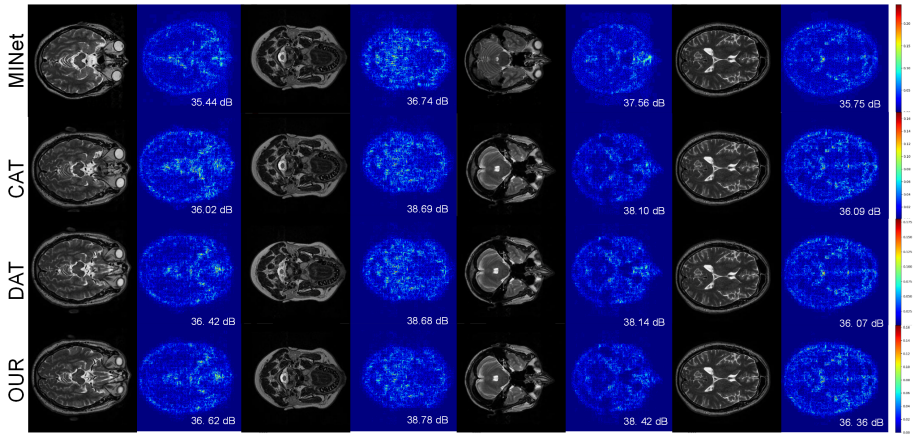

Fig. 4. Reconstructed LR images and their corresponding error map.

$\mathcal{L}_{align\_d}$ contributes the most to final performance in all settings. Combining both alignment losses (Exp. 4) results in even greater improvements, underscoring their complementary effects. The fundamental reason for this performance gain is that $\mathcal{L}_{align\_d}$ constrains the codebook matching process, ensuring matching consistency between LR and HR domains. This reduces reconstruction errors caused by significant discrepancies in matching due to degradation in the LR images. Additionally, the introduction of SAM-generated priors and the spatial cross-attention block further enhance the final performance. This demonstrates the importance of incorporating anatomical priors and focusing on inter-slice information for MRI super-resolution tasks. Overall, each component plays a crucial role in enhancing model capability, with the alignment losses providing the most significant improvements.

To further investigate the impact of $\mathcal{L}_{align\_d}$ on codebook matching, we compute the activations of codebook in test set, considering both HR images and their corresponding 2× and 4× downsampled counterparts. Then, we compare the activation distributions with and without $\mathcal{L}_{align\_d}$. Kullback-Leibler (KL) divergence is employed to quantitatively assess the discrepancy between codebook activations in the HR and LR domains. The results, summarized in Table 3 show that introducing $\mathcal{L}_{align\_d}$ substantially reduces the divergence, indicating improved matching consistency and validating the effectiveness of our approach.

Qualitative study. We visualize the error map of the reconstructed low-resolution images and their corresponding error maps in Fig. 4 to qualitatively verify the effectiveness of our proposed PGASR. We compare our proposed method with the best-performing methods in this study. From the results shown in the figure, it is evident that the reconstruction error of our proposed method is significantly smaller than that of the other two representative methods. This demonstrates the superior performance of PGASR in accurately reconstructing high-quality images with smaller errors. Moreover, the generated anatomical prior shown in Fig. 2 indicates that our method is capable of producing reasonable coarse anatomical structure maps that delineate different MRI regions without requiring additional manual annotation, highlighting the strength and practical utility of our approach.

4 Conclusion

This work addresses the challenge of MRI SR by generating and incorporating anatomical priors. The proposed PGASR framework contains an Anatomical-Conditioned Codebook Generation (ACG) module and a Prior Matching Alignment (PMA) module. The ACG module generates anatomical priors, which are used as conditions to guide codebook generation and preserve structural information. PMA module is designed to enforce probabilistic consistency in codebook index matching across domains and slices, thereby ensuring precise reconstruction of degraded LR images. Experimental validation on the public brain MRI dataset demonstrates the superiority of the proposed PGASR over existing methods, achieving state-of-the-art performance in both fidelity metrics and perceptual quality.

Acknowledgments. This work was supported by CUHK 2022/23 Direct Grant

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

References

1. Q. Lyu, H. Shan, and G. Wang, "Mri super-resolution with ensemble learning and complementary priors," *IEEE Transactions on Computational Imaging*, vol. 6, pp. 615–624, 2020.
2. C.-M. Feng, Y. Yan, H. Fu, L. Chen, and Y. Xu, "Task transformer network for joint mri reconstruction and super-resolution," in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part VI 24*. Springer, 2021, pp. 307–317.
3. C.-M. Feng, H. Fu, S. Yuan, and Y. Xu, "Multi-contrast mri super-resolution via a multi-stage integration network," in *Medical Image Computing and Computer*

Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part VI 24. Springer, 2021, pp. 140–149.
4. H. Yang, Z. Wang, X. Liu, C. Li, J. Xin, and Z. Wang, "Deep learning in medical image super resolution: a review," *Applied Intelligence*, vol. 53, no. 18, pp. 20 891–20 916, 2023.
5. J. Yang, J. Wright, T. S. Huang, and Y. Ma, "Image super-resolution via sparse representation," *IEEE transactions on image processing*, vol. 19, no. 11, pp. 2861–2873, 2010.
6. P. Esser, R. Rombach, and B. Ommer, "Taming transformers for high-resolution image synthesis," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 12 873–12 883.
7. M. Protter, M. Elad, H. Takeda, and P. Milanfar, "Generalizing the nonlocal-means to super-resolution reconstruction," *IEEE Transactions on image processing*, vol. 18, no. 1, pp. 36–51, 2008.
8. C. Chen, X. Shi, Y. Qin, X. Li, X. Han, T. Yang, and S. Guo, "Real-world blind super-resolution via feature matching with implicit high-resolution priors," in *Proceedings of the 30th ACM International Conference on Multimedia*, 2022, pp. 1329–1338.
9. R. Qin, M. Sun, F. Zhang, X. Wen, and B. Wang, "Blind image super-resolution with rich texture-aware codebook," in *Proceedings of the 31st ACM International Conference on Multimedia*, 2023, pp. 676–687.
10. A. D. Gilmore, N. J. Buser, and J. L. Hanson, "Variations in structural mri quality significantly impact commonly used measures of brain anatomy," *Brain informatics*, vol. 8, pp. 1–15, 2021.
11. A. Kirillov, E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, T. Xiao, S. Whitehead, A. C. Berg, W.-Y. Lo *et al.*, "Segment anything," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2023, pp. 4015–4026.
12. Z. Chen, Y. Zhang, J. Gu, L. Kong, X. Yang, and F. Yu, "Dual aggregation transformer for image super-resolution," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2023, pp. 12 312–12 321.
13. J. Zhang, Y. Zhang, J. Gu, Y. Zhang, L. Kong, and X. Yuan, "Accurate image restoration with attention retractable transformer," *arXiv preprint arXiv:2210.01427*, 2022.
14. Z. Chen, Y. Zhang, J. Gu, L. Kong, X. Yuan *et al.*, "Cross aggregation transformer for image restoration," *Advances in Neural Information Processing Systems*, vol. 35, pp. 25 478–25 490, 2022.