Scholar Publishing Group
International Journal of Big Data Intelligent Technology
https://doi.org/10.38007/IJBDIT.2024.050116
ISSN 2790-0932 Vol. 5, Issue 1: 153-160

Super-resolution analysis of CT images based on HAT and self-integrated fusion method

Yunhe Li[1,a], Mei Yang[1,b], Tao Bian[1,c], Wenhui Xia[1,d] and Haitao Wu[2,e*]

[1]Zhaoqing University, Zhaoqing 526060, Guandong, China
[2]Shenzhen Chenzhuo Technology Company, Shenzhen 518055, China
[a]liyunhe@zqu.edu.cn, [b]202308540517@stu.zqu.edu.cn, [c]202408540517@stu.zqu.edu.cn,
[d]202408540530@stu.zqu.edu.cn, [e]frank@cztek.cn
[*]Corresponding author

Keywords: Computed Tomography, Super-resolution, Hybrid Attention Transformer, Self-integrated Fusion Method

Abstract: In this paper, a 4-fold super-resolution analysis scheme for CT images based on HAT (Hybrid Attention Transformer) and self-integrated fusion method is proposed. The scheme realizes the effective capture and fusion of global and local information of CT images by constructing the Dense Residual Transformer Super-Resolution Analysis Architecture (DRTSR), which fuses the self-attention mechanism of Transformer and the local perception capability of Convolutional Neural Network (CNN). Meanwhile, the introduction of the self-integrated fusion method reduces the possible bias of a single model and enhances the stability and accuracy of the results by training multiple HAT models and intelligently fusing them in the testing phase. The experimental results show that the method in this paper demonstrates significant advantages in a number of evaluation indexes, and the generated super-resolution images perform better in terms of naturalness and realism, and the details and sharpness are significantly better than other methods. This work provides higher quality image data support for CT image applications and advances the field of image super-resolution analysis.

1. Introduction

In the field of medical imaging, CT (Computed Tomography) images, as a key diagnostic tool, have an inestimable value in terms of resolution for accurately reflecting the structure of biological tissues and enhancing the precision and efficiency of clinical diagnosis[1]. However, due to the limitations of hardware technology, scanning time considerations, and safety constraints on radiation dose[2], the resolution of conventional CT images is often difficult to meet the standards of high-precision applications, which to a certain extent restricts their wide application and in-depth development in clinical practice[3,4]. Therefore, exploring and developing effective super-resolution reconstruction techniques for CT images to enhance image quality and meet the increasing demand for high accuracy has become an important topic to be addressed[5,6].

In recent years, the rapid development of deep learning technology has revolutionized the field of image super-resolution reconstruction[7]. In particular, super-resolution methods based on the Transformer model have demonstrated excellent capabilities in capturing global contextual information of an image by virtue of its unique self-attention mechanism[8]. The HAT (Hybrid Attention Transformer) model[9], as one of the leaders, realizes the comprehensive capture and deep fusion of global and local information of images by skillfully fusing the self-attention mechanism of Transformer and the local perception ability of Convolutional Neural Networks (CNN), and thus achieves a remarkable Effect.

However, although the HAT model has shown strong potential in image super-resolution tasks, a single model may still have limitations in processing complex and variable CT images[10,11]. To overcome this challenge, this paper further introduces a self-integrated fusion method. This method is based on training multiple HAT models and utilizing the diversity and complementarity among the models for intelligent fusion in the testing stage, thus effectively reducing the possible bias of a single model and significantly enhancing the stability and accuracy of the results. This strategy not only further improves the quality of super-resolution reconstruction, but also provides a more comprehensive and reliable solution for CT image super-resolution analysis.

The aim of this paper is to propose an innovative scheme for 4-fold super-resolution analysis of CT images by deeply studying and exploring the application of HAT model and self-integrated fusion method in super-resolution analysis of CT images. The scheme will fully utilize the advantages of the HAT model in capturing both global and local information and incorporate a self-integrated fusion approach to further enhance the quality of the super-resolution reconstruction. We expect that the research in this paper can provide a new idea and method for the field of super-resolution reconstruction of CT images, promote the technical progress and application development in this field, and provide more accurate, clear and reliable image data support for clinical diagnosis.

In order to achieve this goal, this paper will introduce and analyze the HAT model and the self-integrated fusion method in detail, and construct a super-resolution reconstruction model for CT images based on these two methods. Through a series of well-designed experiments and comparative analysis, we will comprehensively evaluate the effectiveness and superiority of the proposed scheme. We firmly believe that the research and exploration in this paper will bring new breakthroughs and progress in the field of CT image super-resolution analysis.

2. Methods

We present a solution for enhancement utilizing HAT and a self-integrated fusion approach. We constructed the Dense Residual Transformer Super-Resolution Analysis Architecture (DRTSR) to enhance the sensory field by integrating multilevel residuals and dense connections. The dense residual transformer super-resolution analysis architecture (DRTSR) is shown in Fig. 1.

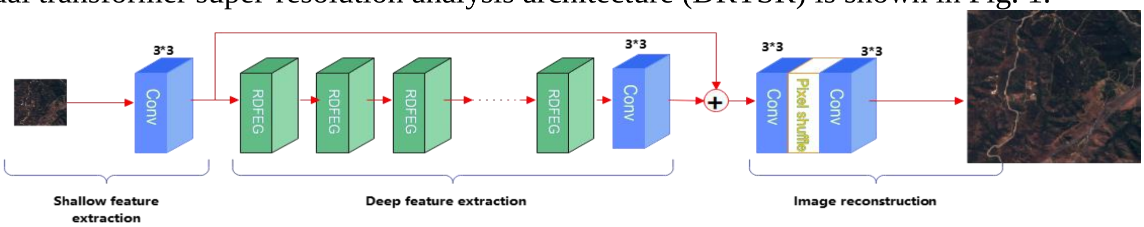

Fig. 1 Schematic diagram of the dense residual transformer super-resolution analysis architecture (DRTSR)

The Dense Residual Transformer Super Resolution Analysis Architecture (DRTSR) consists of three main components: shallow feature extraction, deep feature extraction, and image reconstruction. Shallow feature extraction is first performed using a single 3×3 convolutional kernel, while deep feature extraction is performed using eight residual depth feature extraction groups (RDFEG) paired with a single 3×3 convolution, and image reconstruction is not differentiated between that achieved through the 3×3 convolutional kernel and up-sampling.

The Residual Depth Feature Extraction Group (RDFEG) is realized by a set of Dense Residual Connectivity Blocks (DRCB) connected, and the structure of the RDFEG is shown in Fig. 2.

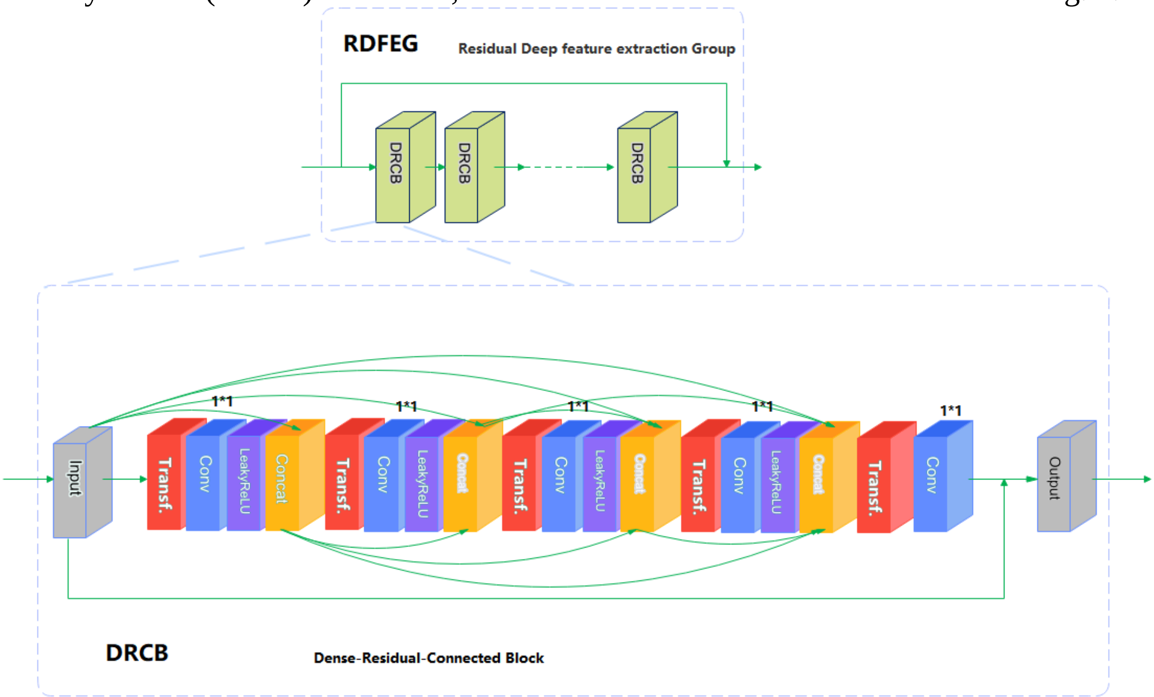

Fig. 2 Schematic diagram of residual depth feature extraction group (RDFEG)

2.1. HAT model construction

The HAT model is the core part of the methodology in this paper, which skillfully combines the advantages of Transformer and CNN, aiming at efficient and accurate 4x super-resolution analysis of CT images. The design of the model fully considers the characteristics of the image super-resolution task, and realizes the effective processing of low-resolution images and the fine reconstruction of high-resolution images through the collaborative work of several key components.

In the initial stage of the HAT model, we use a convolutional layer for the initial processing of the input low-resolution CT images. The purpose of this step is to extract shallow features of the image to provide a basis for subsequent processing. The convolutional layer can effectively capture the basic information such as edges and textures in the image through its local sensing field and parameter sharing characteristics, which lays a solid foundation for the subsequent feature extraction and reconstruction.

In order to further enhance the representation of image features, the HAT model introduces the Transformer's self-attention mechanism. With the multi-head attention module, the model is able to capture long-range dependencies in the image, i.e., correlation information between different regions. This capability is especially important for CT images, which often contain a great deal of complex geographic information and spatial structure. Meanwhile, in order to fully utilize the advantages of CNN in image feature extraction, we combine the convolutional layer of CNN in the HAT model.

The convolutional layer, through its hierarchical feature extraction approach, is able to gradually extract the deeper features in the image, further enhancing the expressive power of the model.

In order to realize the effective fusion of global and local information of images, the HAT model designs a hybrid attention mechanism.The mechanism combines the Transformer's self-attention with the CNN's convolutional attention to capture global and local features in the image, respectively, through parallel processing. Specifically, the self-attention mechanism captures global-scale dependencies by calculating the similarity between different locations in the image, while the convolutional attention mechanism extracts local features in the image through convolutional operations. By fusing the outputs of these two attention mechanisms, the HAT model is able to realize the comprehensive capture and effective utilization of image information.

After extracting the rich image features, the HAT model maps the deep features back to the high-resolution space through the upsampling layer to generate a super-resolution image. The upsampling layer usually employs techniques such as inverse convolution, subpixel convolution, or pixel shuffling to achieve an increase in image resolution. In the HAT model, we have chosen pixel shuffling as the up-sampling method because it reduces the computational complexity and improves the quality of the generated image while maintaining the image details.

In summary, the HAT model achieves efficient super-resolution analysis of CT images through the synergistic work of several components such as shallow feature extraction, deep feature extraction, hybrid attention mechanism and image reconstruction. The model not only integrates the advantages of Transformer and CNN, but also fully considers the characteristics of the image super-resolution task, which provides strong technical support for the application of CT images.

2.2. Self-integrated fusion method

The self-integratied fusion method is an effective technical tool aimed at improving the robustness and stability of the whole system by fusing the predictions of multiple models. In this paper, we apply the self-integrated fusion method to the training and testing process of the HAT model with a view to further improving the performance of 4-fold super-resolution analysis of CT images.

In the training phase, we trained multiple HAT models simultaneously. These models may have subtle differences in initialization parameters, data augmentation methods, or training strategies, which cause them to differ in their learned feature representations and prediction capabilities. By training multiple models, we can obtain an ensemble of models in which each model is capable of making independent super-resolution predictions on input low-resolution CT images.

In the testing phase, we integrate the prediction results from multiple HAT models. Integration can be done by simple averaging, where pixel-level averaging is computed on the super-resolution image output from each model; by more complex weighted averaging, where different weights are assigned to each model based on its performance, and then weighted averaging is performed; or by a voting mechanism, where the predicted values of each pixel are voted on, and the prediction with the highest number of occurrences is selected as the final result.

By fusing the prediction results from multiple models, we can reduce the bias and uncertainty that may be introduced by a single model, thus improving the robustness of the whole system. Even if one model makes a prediction error in some cases, other models may give correct predictions, resulting in a more accurate overall prediction through integration. In practical applications, there are often various noises and variations in the input data. By training multiple models and integrating them, we can better cope with these variations and improve the stability of the system. The integrated prediction results remain relatively stable even if the input data changes somewhat. There may be complementarity between multiple models, that is, they have a different emphasis on the learned feature representation and predictive ability. By integrating their prediction results, we can fully utilize this complementarity to further improve the performance of the whole system. Experimental results show that the self-integrating fusion approach usually leads to significant performance improvement.

In summary, the self-integrated fusion method is an effective technical means by fusing the prediction results of multiple HAT models, we can further improve the robustness, stability and performance of the 4-fold super-resolution analysis of CT images. This method has important value and significance in practical applications.

3. Experiments And Results

3.1. Dataset and experimental setup

In order to validate the effectiveness of the 4-fold super-resolution analysis scheme for CT images based on HAT (Hybrid Attention Transformer) and self-integrated fusion method proposed in this paper, we used the publicly available Chest CT-Scan images dataset[12] for training and testing. The Chest CT-Scan images dataset is a carefully organized dataset of CT images that provides a valuable data resource for super-resolution analysis of CT images.

During the experiments, we preprocessed the dataset in order to improve the generalization ability and robustness of the model. Firstly, the original image was segmented into small chunks of 256x256 pixels so that the model could better learn the local features of the image. Second, data enhancement techniques such as random cropping and flipping are used to further expand the training set and increase the diversity of the data. These preprocessing steps are crucial to improve the training effect and testing performance of the model.

Since the Chest CT-Scan images dataset for super-resolution analysis is already at its highest resolution, and there are no real images that can be tested against the generated images, some commonly used image quality assessment metrics, such as PSNR, SSIM, etc., are no longer applicable in this scenario, and this paper adopts the reference-free image quality assessment (IQA) metrics[13,14], which include NIQE: NIQE is a fully blind image quality assessment model, which is based on a simple and effective statistical model of natural scenes in the spatial domain to establish a “quality-aware” set of statistical features, and is trained using only the measurable deviations from the statistical regularities observed in natural images[15]. The evaluation value of NIQEE can be calculated by the corresponding function niqe in Matlab, and the outputs of the function are all positive real numbers of [0, 100], and all low scores indicate high perceived quality, and high scores indicate low perceived quality. These metrics are able to objectively and fairly evaluate the quality of the generated images and provide us with a strong basis for assessment.

3.2. Analysis of results

The dense residual transformer super-resolution analysis architecture (DRTSR) proposed in this paper is used to do comparative experiments with the current state-of-the-art CT image super-resolution analysis methods BiCubic, EDSR8 RGB, RCAN, and RS-ESRGAN. The evaluation values of NIQE are calculated one by one, and the distribution of evaluation values is plotted as a histogram shown in Fig. 3, and the mean and extreme values according to the evaluation value statistics are shown in Table 1. The experimental results show that compared with existing state-of-the-art methods, the scheme proposed in this paper exhibits significant advantages in several evaluation metrics. Specifically, the NIQE score of this paper's method is much lower than that of other methods, which indicates that the generated super-resolution images perform better in terms of naturalness and realism.The NIQE metric evaluates the quality of an image by measuring its natural statistical properties, and the lower the score indicates that the image is closer to a natural image, so the excellent performance of this paper's method in this metric fully proves the high quality of its generated images.

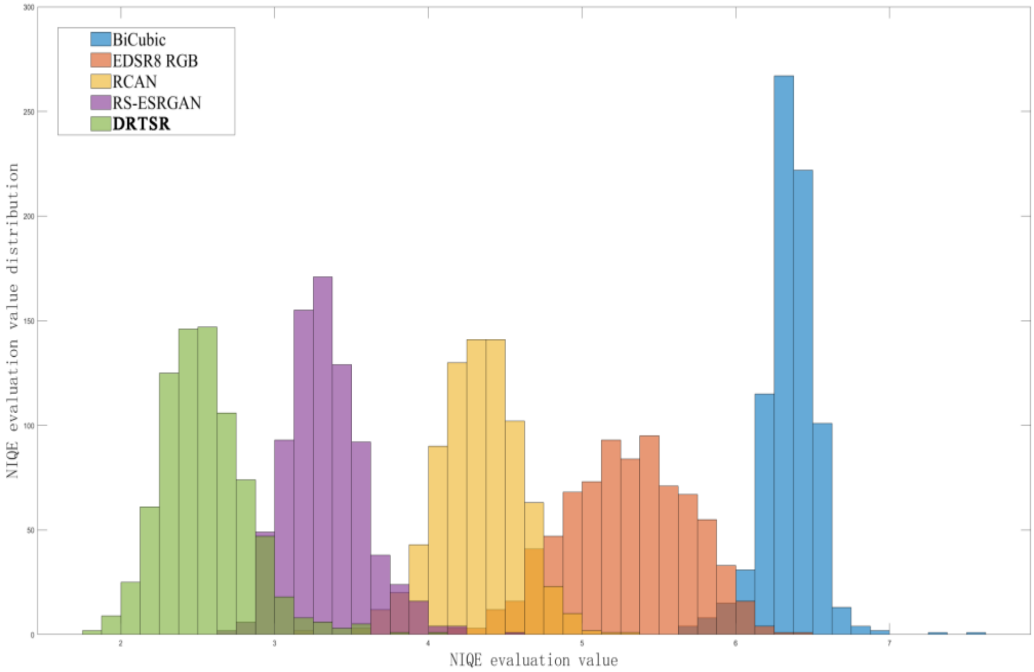

Fig. 3 Distribution of NIQE metrics for experimental results.

Table 1: Statistical data on NIQE assessment values

	Bicubic	**EDSR8-RGB**	**RCAN**	**RS-ESRGAN**	**DRTSR**
NIQE mean	5.852	5.048	4.110	3.745	2.295
NIQE max	6.683	5.285	4.756	5.199	3.453
NIQE min	4.209	3.904	2.730	2.871	1.027

In addition to the objective indicators, we also visualize the advantages of the method in this paper through visual comparisons. Through the comparison of the region images in Fig. 4, it is relatively intuitive to see that the traditional BiCubic method results in the most blurred and smooth images due to the inherent deficiencies of the interpolation algorithm. EDSR8 RGB, RCAN and RS-ESRGAN methods cannot distinguish the noise at the sharpened edges correctly, which leads to blurring of the results, and the demarcation line is clearer in our DRTSR results. Also, it can be seen in the figure that our DRTSR produces very little noise and artifacts compared to existing methods, which indicates that the noise estimated by noise injection is closer to the real noise. Compared with EDSR8 RGB, RCAN and RS-ESRGAN methods, our DRTSR results are clearer and free of ambiguity. The experimental results show that the super-resolution images generated by this paper's method are significantly better than the other methods in terms of both detail and clarity. The method in this paper is able to recover the edge and texture information of the image more accurately, showing higher image quality and stronger detail representation.

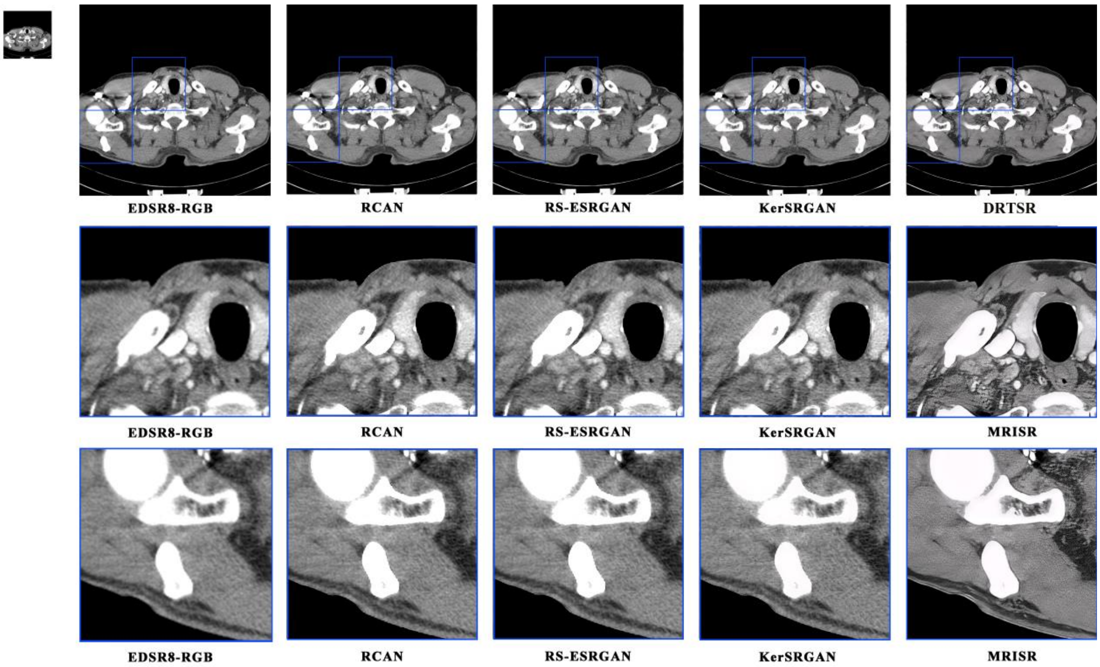

Fig. 4 Comparison of visual effects of generated images

In summary, the 4-fold super-resolution analysis scheme for CT images based on HAT and self-integrated fusion method proposed in this paper achieves excellent performance in a number of evaluation indexes, which fully proves its effectiveness and practicality. The scheme can not only provide higher quality image data support for the application of CT images, but also promote the development and innovation in the field of image super-resolution analysis.

4. Conclusion

In this paper, we propose an innovative 4-fold super-resolution analysis scheme for CT images, DRTSR, which skillfully combines the HAT (Hybrid Attention Transformer) model and the self-integrated fusion method. By introducing the HAT model, we fully utilize the self-attention mechanism of Transformer and the local perception capability of convolutional neural network (CNN) to achieve effective capture and fusion of global and local information of CT images. Meanwhile, through the self-integrated fusion method, we further improve the quality of super-resolution reconstruction, reduce the possible bias of a single model, and enhance the stability and accuracy of the results. The experimental results show that compared with existing state-of-the-art methods, the scheme proposed in this paper exhibits significant advantages in several evaluation metrics. Specifically, the NIQE score of this paper's method is much lower than that of other methods, and the generated super-resolution images perform better in terms of naturalness and realism. In addition, the advantages of this paper's method are also visualized through visual comparisons, where the generated super-resolution images are significantly better than other methods in terms of both detail and clarity, and are able to more accurately recover the edge and texture information of the images. The 4-fold super-resolution analysis scheme for CT images based on HAT and self-integrated fusion methods proposed in this paper not only provides higher quality image data support for CT image applications, but also promotes the development and innovation in the field of image super-resolution analysis. In the future, we will continue to explore more advanced models and methods to further enhance the performance and application scope of image super-resolution analysis.

Funding

This research was funded by the General University Key Field Special Project of Guangdong Province, China (Grant No. 2024ZDZX1004); the General University Key Field Special Project of Guangdong Province, China (Grant No. 2022ZDZX1035)

References

[1] Zhou, S. Kevin, et al. "A review of deep learning in medical imaging: Imaging traits, technology trends, case studies with progress highlights, and future promises." Proceedings of the IEEE 109.5 (2021): 820-838.

[2] Chen, Hongchi, et al. "Deep learning-based algorithms for low-dose CT imaging: A review." European Journal of Radiology (2024): 111355.

[3] Goodfellow, Ian, et al. "Generative adversarial networks." Communications of the ACM 63.11 (2020): 139-144.

[4] Wang X, Yu K, Wu S, et al. Esrgan: Enhanced super-resolution generative adversarial networks[C]//Proceedings of the European conference on computer vision (ECCV) workshops. 2018: 0-0.

[5] Jiang, Yuchen, et al. "Quo vadis artificial intelligence?." Discover Artificial Intelligence 2.1 (2022): 4.

[6] Han, Kai, et al. "Transformer in transformer." Advances in neural information processing systems 34 (2021): 15908-15919.

[7] Liang, Jingyun, et al. "Swinir: Image restoration using swin transformer." Proceedings of the IEEE/CVF international conference on computer vision. 2021.

[8] Liu, Yihao, et al. "Evaluating the generalization ability of super-resolution networks." IEEE Transactions on pattern analysis and machine intelligence (2023).

[9] Chen, Xiangyu, et al. "Hat: Hybrid attention transformer for image restoration." arXiv preprint arXiv:2309.05239 (2023).

[10] Zhu, Lianghui, et al. "Vision mamba: Efficient visual representation learning with bidirectional state space model." arXiv preprint arXiv:2401.09417 (2024).

[11] Demir, Ugur, and Gozde Unal. "Patch-based image inpainting with generative adversarial networks." arXiv preprint arXiv:1803.07422 (2018).

[12] Wen, Long, et al. "A new transfer learning based on VGG-19 network for fault diagnosis." 2019 IEEE 23rd international conference on computer supported cooperative work in design (CSCWD). IEEE, 2019.

[13] Eldarova, E. L. M. I. R. A., V. A. L. E. R. Y. Starovoitov, and K. A. Z. I. Z. A. T. Iskakov. "Comparative analysis of universal methods no reference quality assessment of digital images." (2021).

[14] Mittal, Anish, Anush Krishna Moorthy, and Alan Conrad Bovik. "No-reference image quality assessment in the spatial domain." IEEE Transactions on image processing 21.12 (2012): 4695-4708.

[15] Li, Yunhe, et al. "4× Super‐resolution of unsupervised CT images based on GAN." IET Image Processing 17.8 (2023): 2362-2374.