

### Research Article

## Super-Resolution Swin Transformer and Attention Network for Medical CT Imaging

# Jianhua Hu,<sup>1</sup> Shuzhao Zheng,<sup>1</sup> Bo Wang,<sup>1</sup> Guixiang Luo,<sup>2</sup> WoQing Huang,<sup>1</sup> and Jun Zhang<sup>1</sup>

<sup>1</sup>Computer Engineering Technical College, Guangdong Polytechnic of Science and Technology, Zhuhai Guangdong, China <sup>2</sup>School of Computer Science, Guangdong Polytechnic Normal University, Guangzhou Guangdong, China

Correspondence should be addressed to Jun Zhang; 546949043@qq.com

Received 19 August 2022; Revised 20 September 2022; Accepted 28 September 2022; Published 8 December 2022

Academic Editor: Chen Li

Copyright © 2022 Jianhua Hu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Computerized tomography (CT) is widely used for clinical screening and treatment planning. In this study, we aimed to reduce Xray radiation and achieve high-quality CT imaging by using low-intensity X-rays because CT radiation is damaging to the human body. An innovative vision transformer for medical image super-resolution (SR) is applied to establish a high-definition image target. To achieve this, we proposed a method called swin transformer and attention network (STAN) that uses the swin transformer network, which employs an attention method to overcome the long-range dependency difficulties encountered in CNNs and RNNs to enhance and restore the quality of medical CT images. We adopted the peak signal-to-noise ratio for performance comparison with other mainstream SR reconstruction models used in medical CT imaging. Experimental results revealed that the proposed STAN model yields superior medical CT imaging results than the existing SR techniques based on CNNs. The proposed STAN model employs a self-attention mechanism to more effectively extract critical features and longrange information, hence enhancing the quality of medical CT image reconstruction.

#### 1. Introduction

Computerized tomography (CT) images are used by doctors in clinical practice to judge a patient's condition. Good image quality is crucial for effective and accurate screening and diagnosis of a patient's condition [1]. CT imaging played a vital role in the diagnosis and treatment of COVID-19 [2–4]. CT images are obtained using X-rays. However, X-ray radiation is harmful to the human body. Therefore, to reduce the auxiliary radiation, the X-ray intensity is reduced during operation, resulting in low resolution and blurring of CT images. Therefore, how to obtain high-definition medical CT images through superresolution (SR) is an important research topic. For high-quality medical CT images, plenty of deep learning-(DL-) based SR techniques have been proposed [4–6].

Convolutional neural networks (CNNs) have been used to accomplish SR tasks. Initially, the SRCNN network and CNN were used for performing high-resolution reconstruction tasks [7]. This is the earliest reconstruction from lowresolution to high-resolution images by using CNNs and point-to-point nonlinear feature mapping and reconstruction. Currently, DL is widely used in SR applications [8–10]. Shan et al. [11] improved the initial CNN-based SR method by introducing residual learning and attention mechanism.

Powerful reconstruction algorithms have been proposed to improve SR capability. In 2016, FSRCNN [12] was proposed to improve the SRCNN model, and upsampling was performed to increase the running speed. Many CNNbased SR algorithms have been proposed to improve residual learning, attention mechanism, model depth, speed, complexity reduction, and SR performance [13–22].

CNNs are used in the mainstream medical CT image SR algorithms because they provide a very high-performance advantage for the image domain. However, CNNs cannot realize long-range feature extraction. The transformer is mostly employed in the audio industry, but it has been, recently, employed for SR as a replacement for CNNs [23, 24] because the transformer can support long-range feature extraction by using a self-attention (SA) mechanism and yields very good performance in the image domain. Transformer models in the field of medical imaging have been extensively studied [25]. The transformer network and the shared attention approach to limit feature extraction improve the SR performance. Nevertheless, few studies have used the transformer network to improve the SR of medical images. As such, in this study, we attempted to use the transformer network for SR reconstruction of medical CT images.

To reconstruct medical CT images, in this study, we developed the swin transformer and attention network (STAN) model. The main advantage of STAN is that it can learn feature information better. STAN consists of three types of blocks: the low-frequency feature extraction block, deep feature extraction block (including attention transformer blocks (ATBs)), and high-resolution image reconstruction block.

To preserve the low-frequency information, the lowfrequency feature extraction block is directly connected to the reconstructed model. The deep feature extraction module mainly consists of ATBs. To extract image edge and texture information, a shift window size is used, which reduces resource consumption. Finally, in the high-resolution image reconstruction block, the features of the first two models are first obtained, multilayer feature fusion is performed, and finally, lowresolution to high-resolution reconstruction is realized.

The main contributions of this study are as follows:

- (1) A swin transformer is proposed in this paper for an SR network of medical CT images. The use of the attention mechanism improves the network's ability to extract features and edge and flat area information from medical CT images and reconstruct highquality CT images
- (2) We developed a low-frequency extraction module with an attention mechanism to capture the longrange dependency feature of the image
- (3) To handle long-range dependency images, we used a shift window mechanism, overcoming the traditional transformation limitation of dividing the input image into fixed-size patches

#### 2. Related Work

The traditional SR algorithm uses a bicubic interpolation algorithm to upsample an image and has the disadvantages of losing details and blurring the image. Therefore, neural networks have been employed for SR. The transformer network can further improve the performance of a traditional CNN. With the development of SR, many scholars have applied the SR technology to improve the clarity and reliability of medical CT images by employing the following three approaches:

(1) Obtaining SR images by using CNNs: CNNs are mainly used to perform transformations between images of different resolutions (e.g., LR image to HR image). Due to the different characteristics of the image, different image scaling methods need to be used to recover different image details. Therefore, nonlinear mapping is performed to recover the lost high-frequency details. CNNs are widely used to reconstruct high-quality images and realize SR through dense connection convolution, multichannel networks, and symmetric jump connections [26, 27]

- (2) The use of transformer networks in the field of image applications: transformer networks are generally used in the audio field. Because of their local attention mechanism and long-term compliance, transformer networks are highly suitable for image feature extraction. Therefore, the transformer [28, 29] is widely used in the field of image processing because of its ability to better access information and integrate the CNN and transformer. Pan et al. [23] proposed a high-quality reconstruction transformer to capture image global features for medical CT image reconstruction
- (3) The use of SR in the field of medical CT imaging: DL technology is extensively employed for medical CT imaging [30–32]. Many scholars have applied SR technology to the medical field [33–35]. SR technology is used to reconstruct high-definition images for the characteristics of medical images, which can effectively improve image quality and reduce X-ray radiation to the human body

In this study, we designed the STAN model to reconstruct medical CT images. In addition, we introduced a self-information mechanism in the network model to enable updates to be performed on long-range information; moreover, the medical image quality smoothing area enables better image quality.

#### 3. Methods

The architecture of the medical CT image performance enhancement network is presented here. For the SR reconstruction of medical CT images, a transformer and an attention network are employed. To improve the extraction of low-frequency and high-frequency medical feature information, we designed the STAN model. We used a transformer network instead of traditional CNNs to considerably increase the quality of medical CT images and edge information. The proposed system comprises three types of blocks: low-frequency feature extraction block, deep feature extraction block, and high-resolution image reconstruction block.

3.1. Network Architecture. The structure of the proposed STAN model is illustrated in Figure 1. STAN employs an efficient long-range attention transformer network for reconstructing high-resolution images from low-resolution medical CT images. STAN includes the low-frequency feature extraction block (for the extraction of flat area image information), deep feature extraction block (including six ATBs), and high-resolution CT image reconstruction block (for the feature extraction of image edge information). Low-resolution images are inputted into the STAN model.



FIGURE 1: The architecture of the proposed STAN model. The STAN includes the low-frequency feature extraction block (for the extraction of flat area image information), deep feature extraction block (including six ATBs), and high-resolution CT image reconstruction block (for the feature extraction of image edge information).

**Input:** low-resolution image input *I*<sub>LO</sub> Output: high-resolution image output I<sub>RHO</sub> 1:  $F_0 = H_{LF} * I_{LQ}$ 2: For  $i = 1 \longrightarrow M do//i$  denotes the *i*-th ATB, here *M* is 6 3:  $//F_i = H_{ATB}(F_i)$ For  $j = 1, 2, \dots, L$  do//j denotes the j-th STL layer, L is 6 4:  $F_{i,j} = H_{\text{STL}i,j} \left( F_{i,j} - 1 \right)$ 5: 6: End for  $F_{i,\text{out}} = H_{\text{CONV}i}(F_{i,L}) + F_{i,0}$ 7: 8: End for 9:  $F_{\text{DF}} = H_{\text{CONV}}(F_M)$ 10:  $I_{\rm RHQ} = H_{\rm REC} (F_0 + F_{\rm DF})$ 11: Return I<sub>RHO</sub>//high-resolution image output

ALGORITHM 1: The implementation steps of our STAN for medical superresolution.

The low-frequency feature extraction block extracts the lowfrequency feature information from medical CT images by using multilayer CNNs. The deep feature extraction block employs the self-attention mechanism transformer network to extract the edge information of medical CT images, and multichannel image information is obtained by adding to the previous network. The high-resolution CT image reconstruction block fuses the low-frequency and high-frequency information; in addition, it extracts the characteristic data of multiple channels and upsamples the image to obtain an SR medical CT image.

The proposed STAN algorithm is shown below. The low-resolution input image is  $I_{LQ}$ . Transform  $H_{LF}$  is a low-frequency feature extraction block. In the deep feature extraction block, there are *M* ATBs, and each ATB has *L* STLs and a convolution operation. After the deep feature extraction block, the high-quality picture is reconstructed through high-resolution CT image reconstruction.

*3.2. Low-Frequency Feature Extraction Block.* As shown in Figure 2, the low-frequency feature extraction block realizes low-frequency information extraction and includes three

layers. A low-resolution image is input in. After feature extraction by using  $3 \times 3$  convolution operations, fine feature extraction is performed using  $1 \times 1$  convolution. Finally, the low-frequency information extraction output of the current block is obtained using a convolution kernel size of  $3 \times 3$ .

A low-resolution image is input as  $I_{LQ}$ . Then, two  $3 \times 3$  and one  $1 \times 1$  convolution layer are used to obtain the low-frequency feature output as follows:

$$F_0 = H_{\rm LF} \left( I_{\rm LQ} \right). \tag{1}$$

This module uses a multilayer network to better accomplish the extraction of low-frequency information.

3.3. Deep Feature Extraction Block. Deep features  $F_{\text{DF}} \in \mathbb{R}^{H \times W \times C}$  are extracted from the low-frequency feature output  $F_0$  as follows:

$$F_{\rm DF} = H_{\rm DF}(F_0),\tag{2}$$

where comprises six ATBs. The composition and principle of the ATB are described in detail here.



FIGURE 3: Attention transformer block.

As shown in Figure 3, ATBs are composed of swin transformer layers (STLs) and convolutional layers with selfawareness functions. The STL is the base component of the ATB. The base network comprises multiple STLs and ends with a convolutional layer to form the ATB. In this study, the number of STLs in an ATB was set as 6 to achieve a balance between extraction performance and model complexity.

 $F_{i,0}$  indicates the *i*-th ATB. Information features  $F_{i,1}$ ,  $F_{i,2}$ , ...,  $F_{i,L}$  are extracted by the ATB layers as follows:

$$F_{i,j} = H_{\text{STL}_{i,j}}(F_{i,j-1}), j = 1, 2, \cdots, L,$$
(3)

where  $H_{\text{STL}_{i,j}}(\cdot)$  is the *i*-th ATB and *J* denotes the *j*-th STL. This design offers two advantages: spatial variation convolution and residual connection reconstruction module.

The output of ATB can be formulated as

$$F_{i,\text{out}} = H_{\text{CONV}}(F_{i,L} + F_{i,0}), \tag{4}$$

where  $H_{\text{CONV}_i}(\cdot)$  is the *i*-th ATB swin transformer.

The STL enables the self-attentive mechanism through the transformer layer. Its most important feature is the use of local attention and shift window mechanism. By the size  $H \times W \times C$ , the ATB splits the input into nonoverlapping  $M \times M$  local windows. In this manner, the input size is reshaped into the  $(HW/M^2) \times M^2 \times C$  feature, where  $HW/M^2$  is the number of windows.

The STL consists of three components: layer specification (LN) layer (used for regularization), multicontrol head SA (MCSA) layer, and multilayer control perceptron (MLCP) layer. The MLCP layer is composed of two completely connected neural networks, and feature extraction is performed between them through nonlinear transformation. The LN layer is added before the MCSA and MLCP layers, and then, the residuals are used to connect the two modules. The process is as follows:

$$X = MCSA(LN(X)) + X,$$
 (5)

$$X = \mathrm{MCLP}(LN(X)) + X.$$
(6)

*3.4. High-Resolution Image Reconstruction Block.* The low-frequency feature extraction of high-quality images is performed from medical CT images according to the convergence:

$$I_{\rm RHQ} = H_{\rm REC}(F_0 + F_{\rm DF}),\tag{7}$$

where  $H_{\text{REC}}(\cdot)$  indicates the reconstruction model. The low-frequency information mainly includes low frequencies, while the deep features are used to repair the missing high frequencies. Sloshing inverter circuits are used to transmit low-frequency information to the medical CT image



FIGURE 4: High-resolution image reconstruction blocks.

reconstruction module through a high skip connection and help the deep-level feature collection module to focus on high-frequency information.

The high-resolution image reconstruction block (Figure 4) comprises a 64-channel CNN with channel size H/2, w/2, 64. The 64-channel feature map output is obtained using the pixel shuffle upsampling method. Finally, a 3-channel CNN is used to generate the high-definition image output.

The primary function of pixel shuffle is to convert the multichannel feature map r \* r into size of w \* r and h \* r (e.g., the original feature map size is  $4 \times 128 \times 128$ , which is then adjusted to size  $1 * (128 \times 2) * (128 \times 2)$ ), where *r* is the upsampling factor, which is the magnification of the image.

#### 4. Results

We evaluated the performance of the proposed model on open-source datasets and evaluated the image quality by using the peak signal-to-noise ratio (PSNR) metric. Compared with other advanced SR methods, the proposed model offers obvious performance advantages.

4.1. Dataset. We used the largest medical CT medical image dataset, DeepLesion [36], for training and testing the model. This dataset not only includes key CT slices containing the important lesions but also provides the three-dimensional context (additional slices of 30 mm above and below the key slices). The size of the dataset is 221 GB. Because of the huge amount of data, 11,500 high-quality CT images were randomly selected and divided into three parts. The majority of the images were used for training (10,000), and the remaining were used for verification (1000) and testing (500). This dataset consists of the original image and the downsampled image through bicubic interpolation by using the function torchvision.transforms.resize() in the PyTorch library. The source HR medical CT image was reduced to a LR image as the input data, and the original HR medical CT medical image was used as the data label to be used as the input dataset of the DL neural network for training. For the sake of accuracy of model training, the training set was added through data enhancement to improve the generalization ability.

4.2. Implementation Details. The three-channel (RGB) pixels of the input image and the original data were linearly reduced to obtain the LR image, and the original data were used as label data and inputted into the network. Six ATBs were used. The sliding window size of each transform net-

ATB number		
Number of ATBs	X4	PSNR
1		23.3
2		25.78
4		28.967
6		31.34
8		31.90
10		32.02

TABLE 1: Ablation study on ATB design.

work was set as 8, and the patch size corresponding to the LR image resolution was 48.

Adam optimizer was adopted with two improvements: gradient sliding average and bias correction. The learning rate decayed with each update factor decay set as 0.999, and the initial learning rate was  $2 \times 10^{-4}$ . The pixel shuffle method was used for image upsampling.

4.3. Evaluation Index. We evaluated the reconstructed SR images by using two methods: subjective evaluation and objective evaluation. Many factors influence subjective evaluation, and the reconstructed SR images are evaluated mainly based on human visual perception. In this study, the PSNR was measured as the objective evaluation metric to study the performance of high-resolution restoration networks for medical CT images. To demonstrate the superiority of the proposed model visually, we calculated the PSNR values of the SR images generated using the proposed method and other methods and compared them.

PSNR is an objective criterion for evaluating images. The calculation method is as follows:

$$MSE = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} \|J(i,j) - L(i,j)\|^2.$$
(8)

Then, PSNR can be obtained as follows:

$$PSNR = 10 \log_{10} \frac{MaxValue^2}{MSE} = 10 \log_{10} \frac{2^{bits} - 1}{MSE}, \quad (9)$$

where *J* and *L* are the two pixel values and the size of the image is  $m \times n$ . The greater the PSNR, the better the medical CT image effect, and vice versa.

4.4. Ablation Study. To better understand how STAN performs SR in medical CT images, a comprehensive ablation study on ATBs was performed to evaluate the role of key



FIGURE 5: Relationship between PSNR and number of ATBs in STAN.

parts of the proposal STAN model, as well as the degree of depth and the choice of shared attention mechanism.

As can be seen in Table 1, we studied the effect of the removal and addition of ATB modules on the performance of the medical CT image reconstruction network. To analyze the effect of the low-frequency feature extraction block and ATB on the performance of the STAN model, we conducted ablation experiments by using different numbers of modules and studied their corresponding PSNR performance under the ×4 scaling condition. The number of ATBs affects the PSNR, i.e., the higher the number of ATBs, the higher the PSNR.

As shown in Figure 5, we studied the relationship between the number of ATBs and the PSNR performance on DeepLesion for image SR ( $\times$ 4). To obtain a relatively lightweight model, the number of ATBs was selected as 6, and the number of convolutional layers was 3 in the final test performance experiment.

4.5. Analysis of Experimental Results. Network optimization was performed. The performance comparison results in terms of PSNR with  $\times 2$  and  $\times 4$  scale factors are presented in Table 2. We analyzed different algorithms on the DeepLesion testing set. Compared with the bicubic method, the PSNR of STAN improved by 9.58 and 13.36 dB when the scale factor was  $\times 2$  and  $\times 4$ , respectively. Compared with the method using the DL neural network, the PSNR of STAN improved by 3.81 and 3.56 dB when the scale factor was  $\times 2$  and  $\times 4$ , respectively.

As can be seen in Figure 6, the bicubic reconstruction of medical CT images yielded the worst effect and the lowest PSNR. The SR algorithm based on DL performed better than the algorithm based on interpolation. The STAN model based on transform networks proposed in this paper performed relatively better than the CNN-based SR method

TABLE 2: PSNR (dB) values for the proposed STAN model.

×2	×4
23.32	21.76
33.17	27.78
34.56	29.65
34.85	29.90
34.96	30.24
34.99	30.35
36.22	31.11
36.98	31.34
	×2 23.32 33.17 34.56 34.85 34.96 34.99 36.22 36.98

by 0.76 and 0.23 dB when the scale factor was  $\times 2$  and  $\times 4$ , respectively.

Thus, the proposed STAN model exhibited superior performance to the CNN-based SR method, demonstrating that the transformer network yields obvious performance advantages in medical CT imaging.

By using different algorithms, the medical CT image was reconstructed with multiple resolutions. The results of different algorithms on the DeepLesion are shown in Figure 6.

#### 5. Conclusions

For SR of medical CT images, we proposed an improved STAN model that uses the SA mechanism for feature extraction and solves the long-range dependency problem encountered in CNNs and RNNs. In addition, it can obtain more important feature information. In STAN, nonoverlapping feature values are computed using different window sizes, and feature extraction is performed using a shared-attention mechanism.

We experimentally demonstrated the SR effectiveness of the proposed STAN model in medical CT images. We used the PSNR metric for performance comparison. The results



FIGURE 6: Medical CT image testing using different algorithms on the DeepLesion dataset.

revealed that the PSNR of the proposed STAN model is much better than that of the CNN SR method. The use of the SA mechanism in STAN yields clearer reconstruction results, and the reconstruction effect in the low-frequency regions of medical CT images is better. However, medical imaging may generate image noise due to the influence of hardware equipment and the external environment. As such, the next step is to denoise medical CT images in the SR process.

#### **Data Availability**

The medical CT medical image data used to support the findings of this study have been deposited in the https:// nihcc.app.box.com/v/DeepLesion repository. This is an open source medical open source dataset, you can download and use it freely.

#### **Conflicts of Interest**

The authors declare that they have no conflicts of interest.

#### Acknowledgments

The authors acknowledge the Foundation for 2022 Basic and Applied Basic Research Project of Guangzhou Basic Research Plan (research on video compression algorithm based on dual neural network, Grant 202201011753), the 2020 University Industry University Research Innovation Fund New Generation Information Technology Innovation Project of China (key project, Grant 2020ITA03004), the 2021 characteristic innovation project of the Department of Education of Guangdong Province (2021KTSCX217), the Research on Classified and Accurate Training of Higher Vocational IT Talents based on Education Big Data Under the Background of Enrolment Expansion (Grant 2021GXJK714), and the Innovative Research Team in Universities of Guangdong Province of China (Grant 2021KCXTD079).

#### References

D. J. Brenner, C. D. Elliston, E. J. Hall, and W. E. Berdon, "Estimated risks of radiation-induced fatal cancer from pediatric CT," *American Journal of Roentgenology*, vol. 176, no. 2, pp. 289–296, 2001.

- [2] M. Bakator and D. Radosav, "Deep learning and medical diagnosis: a review of literature," *Multimodal Technologies* and Interaction, vol. 2, no. 3, p. 47, 2018.
- [3] C. Long, H. Xu, Q. Shen et al., "Diagnosis of the coronavirus disease (COVID-19): rRT-PCR or CT?," *European Journal of Radiology*, vol. 126, article 108961, 2020.
- [4] M. M. Rahaman, C. Li, Y. Yao et al., "Identification of COVID-19 samples from chest X-ray images using deep learning: a comparison of transfer learning approaches," *Journal of X-Ray Science and Technology*, vol. 28, no. 5, pp. 821–839, 2020.
- [5] Z. Chen, X. Guo, P. Y. Woo, and Y. Yuan, "Super-resolution enhanced medical image diagnosis with sample affinity interaction," *IEEE Transactions on Medical Imaging*, vol. 40, no. 5, pp. 1377–1389, 2021.
- [6] M. I. Georgescu, R. T. Ionescu, A. I. Miron et al., "Multimodal multi-head convolutional attention with various kernel sizes for medical image super-resolution," 2022, https://arxiv.org/ abs/2204.04218.
- [7] J. Kim, J. K. Lee, and K. M. Lee, "Accurate image superresolution using very deep convolutional networks," *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1646–1654, 2016.
- [8] L. S. Passarella, S. Mahajan, A. Pal, and M. R. Norman, "Reconstructing high resolution ESM data through a novel fast super resolution convolutional neural network (FSRCNN)," *Geophysical Research Letters*, vol. 49, no. 4, 2022.
- [9] X. Wang, K. Yu, S. Wu et al., "Enhanced deep residual networks for single image super-resolution," *Proceedings of the European conference on computer vision (ECCV) workshops*, 2018.
- [10] I. Lodha, L. Kolur, K. Krishnan, K. Dheenadayalan, D. Sitaram, and S. Nandi, "Cost-optimized video transfer using real-time super resolution convolutional neural networks," in 5th Joint International Conference on Data Science & Management of Data (9th ACM IKDD CODS and 27th COMAD), Bangalore, India, 2022.
- [11] L. Shan, X. Bai, C. Liu, Y. Feng, Y. Liu, and Y. Qi, "Super-resolution reconstruction of digital rock CT images based on residual attention mechanism," *Advances in Geo-Energy Research*, vol. 6, no. 2, pp. 157–168, 2022.
- [12] C. Dong, C. C. Loy, and X. Tang, "Accelerating the superresolution convolutional neural network," in *Proceedings of the European Conference on Computer Vision (ECCV 2016)*, pp. 391–407, Amsterdam, The Netherlands, 2016.
- [13] M. Ta and B. Ylmaz, "Super resolution convolutional neural network based pre-processing for automatic polyp detection in colonoscopy images," *Computers & Electrical Engineering*, vol. 90, no. 12, article 106959, 2021.

- [14] A. Deshpande, V. V. Estrela, and P. Patavardhan, "The DCT-CNN-ResNet50 architecture to classify brain tumors with super-resolution, convolutional neural network, and the ResNet50," *Neuroscience Informatics*, vol. 1, no. 4, article 100013, 2021.
- [15] M. O. Yaroshenko, A. Y. Varfolomieiev, and P. O. Yaganov, "Hierarchical convolutional neural network for infrared image super-resolution," *Microsystems Electronics and Acoustics*, vol. 26, no. 1, 2021.
- [16] D. Qiu, L. Zheng, J. Zhu, and D. Huang, "Multiple improved residual networks for medical image super-resolution," *Future Generation Computer Systems*, vol. 116, pp. 200–208, 2021.
- [17] Q. Chang, X. Jia, C. Lu, and J. Ye, "Multi-Attention Residual Network for Image Super Resolution," *International Journal* of Pattern Recognition and Artificial Intelligence, vol. 36, no. 8, article 2254009, 2022.
- [18] K. G. Brown, S. C. Waggener, A. D. Redfern, and K. Hoyt, "Faster super-resolution ultrasound imaging with a deep learning model for tissue decluttering and contrast agent localization," *Biomedical Physics & Engineering Express*, vol. 7, no. 6, article 065035, 2021.
- [19] X. Wang, K. Yu, S. Wu et al., "Esrgan: enhanced superresolution generative adversarial networks," in *European Conference on Computer Vision Workshops*, pp. 701–710, Munich, Germany, 2018.
- [20] J. Zhang, X. Zou, L. D. Kuang, J. Wang, R. S. Sherratt, and X. Yu, "CCTSDB 2021: a more comprehensive traffic sign detection benchmark," *Human-centric Computing and Information Sciences*, vol. 12, 2022.
- [21] Y. Chen, L. Liu, V. Phonevilay et al., "Image super-resolution reconstruction based on feature map attention mechanism," *Applied Intelligence*, vol. 51, no. 7, pp. 4367–4380, 2021.
- [22] J. Zhang, C. Li, S. Kosov et al., "LCU-Net: a novel low-cost U-Net for environmental microorganism image segmentation," *Pattern Recognition*, vol. 115, p. 107885, 2021.
- [23] J. Pan, H. Zhang, W. Wu, Z. Gao, and W. Wu, "Multi-domain integrative Swin transformer network for sparse-view tomographic reconstruction," *Patterns*, vol. 3, no. 6, article 100498, 2022.
- [24] B. Zhou, J. Schlemper, N. Dey et al., "DSFormer: a dualdomain self-supervised transformer for accelerated multicontrast MRI reconstruction," 2022, https://arxiv.org/abs/ 2201.10776.
- [25] F. Shamshad, S. Khan, S. W. Zamir et al., "Transformers in medical imaging: a survey," 2022, https://arxiv.org/abs/2201 .09873.
- [26] Y. Zou, F. Xiao, L. Zhang, Q. Chen, B. Wang, and Y. Hu, "Image super-resolution using convolutional neural network with symmetric skip connections," in *Fourth International Conference on Photonics and Optical Engineering*, vol. 11761, pp. 283–288, Xi'an, China, 2021 Jan 15.
- [27] Y. Wang, S. Zeng, X. Wang, and J. Wang, "Super-resolution multi-focus image fusion based on convolutional neural network," *Journal of Physics: Conference Series*, vol. 1885, no. 2, article 022011, 2021.
- [28] Z. Liu, Y. Lin, Y. Cao et al., "Swin transformer: hierarchical vision transformer using shifted windows," 2021, https:// arxiv.org/abs/2103.14030.
- [29] H. Chen, C. Li, X. Li et al., "Gashis-transformer: a multi-scale visual transformer approach for gastric histopathology image classification," 2021, https://arxiv.org/abs/2104.14528.

- [30] R. Kumar, A. A. Khan, J. Kumar et al., "Blockchain-federatedlearning and deep learning models for covid-19 detection using ct imaging," *IEEE Sensors Journal*, vol. 21, no. 14, pp. 16301–16314, 2021.
- [31] K. He, X. Liu, M. Li, X. Li, H. Yang, and H. Zhang, "Noninvasive KRAS mutation estimation in colorectal cancer using a deep learning method based on CT imaging," *BMC medical imaging*, vol. 20, no. 1, p. 59, 2020.
- [32] W. Wu, D. Hu, C. Niu et al., "Deep learning based spectral CT imaging," *Neural Networks*, vol. 144, pp. 342–358, 2021.
- [33] R. Gupta, A. Sharma, and A. Kumar, "Super-resolution using GANs for medical imaging," *Proceedia Computer Science*, vol. 173, pp. 28–35, 2020.
- [34] J. Zhu, C. Tan, J. Yang, G. Yang, and P. Lio, "Arbitrary scale super-resolution for medical images," *International Journal* of Neural Systems, vol. 31, no. 10, article 2150037, 2021.
- [35] J. Zhang, L. R. Gong, K. Yu, X. Qi, Z. Wen, and Q. Hua, "3D reconstruction for super-resolution CT images in the Internet of health things using deep learning," *IEEE Access*, vol. 8, pp. 121513–121525, 2020.
- [36] Y. Ke, X. Wang, L. Le, and R. M. Summers, "DeepLesion: automated mining of large-scale lesion annotations and universal lesion detection with deep learning," *Journal of medical imaging*, vol. 5, no. 3, article 036501, 2018.
- [37] C. Dong, C. C. Loy, K. He, and X. Tang, "Image superresolution using deep convolutional networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 2, pp. 295–307, 2016.
- [38] Y. Tai, J. Yang, and X. Liu, "Image super-resolution via deep recursive residual network," in *Proceedings of the IEEE conference on computer vision and pattern recognition.*, pp. 3147– 3155, Honolulu, Hawaii, 2017.
- [39] B. Lim, S. Son, H. Kim, S. Nah, and K. Mu Lee, "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, Honolulu, Hawaii, 2017.
- [40] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu, "Residual dense network for image super-resolution," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2472–2481, Salt Lake City, Utah, 2018.
- [41] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu, "Image super-resolution using very deep residual channel attention networks," in *Proceedings of the European conference on computer vision (ECCV)*, pp. 286–301, Munich, Germany, 2018.
- [42] J. Lee and K. H. Jin, "Local texture estimator for implicit representation function," in *Proceedings of the IEEE/CVF Conference* on Computer Vision and Pattern Recognition, pp. 1929–1938, Ernest N. Moral Convention Center, New Orleans, 2022.