Transactions on Computer Science and Intelligent Systems Research
ISSN: 2960-1800, eISSN: 2960-2238 | Volume 5
AIDML 2024

Optimization of Super-Resolution Features via Deep Learning Techniques

Yongkang Hu [1, *], Chuyue Jing [2], Ziyuan Li [3] and Xiao Liu [4]

[1] Academy of automation, Central South university, Changsha, China

[2] The School of Materials Science and Engineering (SMSE), Shanghai Jiao Tong University, Shanghai, China

[3] The School of Materials Science and Engineering (SMSE), Chang'an University, Xi'an, China

[4] School of Electronics and Information, Beijing Institute of Technology, Beijing, China

* Corresponding Author Email: 8213210221@csu.edu.cn

Abstract. This manuscript delves into the pivotal role of feature optimization within the realm of deep learning-driven super-resolution techniques, a topic of considerable importance across diverse domains such as medical imaging and electron microscopy. The paper confronts the inherent challenges posed by deep learning in super-resolution tasks, including issues related to complexity, computational burden, and the paucity of model generalizability. It underscores the significance of feature optimization strategies, particularly the application of time-to-frequency transformation techniques such as wavelet and Fourier transforms, as well as attention mechanisms, in enhancing model efficiency and curtailing the demand on computational resources. A central theme of the discourse is the efficacy of amalgamating time-frequency separation methods — notably wavelet and Fourier transforms — with attention mechanisms in super-resolution endeavors. This integrated approach, termed multi-scale feature fusion, is lauded for its capacity to augment the model's attentiveness to high-frequency details, which are crucial in super-resolution. This fusion strategy not only optimizes the allocation of computational resources but also elevates learning performance.

Keywords: Super-Resolution, deep learning, feature optimization, time-to-frequency transformation, wavelet transform.

1. Introduction

In contemporary research, the integration of advanced deep-learning techniques in super-resolution methodologies is proving indispensable in diverse fields, including medical imaging and electron microscopy image processing. Deep learning-based super-resolution (DL-based SR) approaches are rapidly gaining traction due to their unparalleled efficacy [1-4]. As such, the focus of this paper is exclusively on feature optimization within deep-learning models dedicated to super-resolution tasks.

However, implementing deep learning in super-resolution is not without its challenges, predominantly in terms of computational complexity and the issue of model generalization. To address these concerns, a plethora of methodologies centering on feature optimization has emerged. Notably, the application of time-to-frequency transformations, including wavelet and Fourier transforms, has become a common practice for efficient feature extraction [5-8]. Furthermore, the integration of attention mechanisms into learning architectures significantly enhances the model's ability to selectively concentrate on pertinent details within feature maps. Evidence suggests that these feature optimization strategies can markedly mitigate computational complexity and resource consumption [9-12]. This paper endeavors to scrutinize these feature optimization tactics within super-resolution endeavors, forecasting potential challenges and offering insights into the evolving role of feature optimization in the next generation of deep-learning-based super-resolution technologies.

2. Definition of image feature optimization method for Super- Resolution

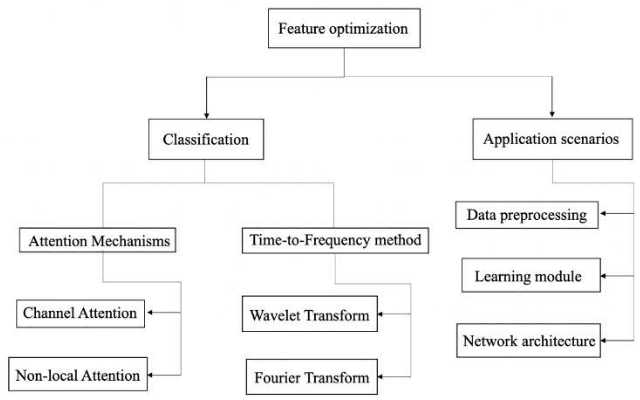

Fig 1. Taxonomy of related work (Photo/Picture credit: Original).

In the realm of feature optimization, methodologies predominantly bifurcate into traditional approaches, distinguishing time and frequency domains, and strategies leveraging attention mechanisms. As shown in Figure 1. Predominantly, the separation of time-frequency functions hinges on the use of wavelet transforms and Fourier transforms. While alternatives such as the Gabor Transform and Stockwell Transform exist, their adoption is less frequent due to their elevated computational demands. The transmutation from the time domain to the frequency domain facilitates the extraction of salient frequency information from images. This transition is adept at capturing frequency characteristics of signals more effectively than time domain analysis, thereby enabling deep learning models to proficiently learn and extrapolate periodic and cyclical variations in signals [13]. Moreover, frequency domain representations, often less redundant than their temporal counterparts, augment neural network learning. Notably, such representations exhibit an inherent robustness to specific noise types, enhancing the resilience of neural networks against noise interference during learning processes [14]. Concurrently, attention mechanisms have gained traction in refining the efficacy of deep learning models. By directing focus towards critical details within feature maps, these mechanisms accentuate key object attributes like boundaries and textures, particularly in challenges like super-resolution. Simultaneously, they mitigate the influence of trivial elements (e.g., noise or low-frequency details), thereby significantly bolstering model learning efficiency and reducing computational strain. Although attention mechanisms and time-frequency transformation techniques are intrinsically distinct, their synergistic integration can yield exceptional outcomes in super-resolution tasks [15]. The conversion from time to frequency domain plays a crucial role in distilling frequency information from images, allowing for the discernment of high-frequency details such as textures and edges, which are pivotal in super-resolution. The amalgamation of attention mechanisms with time-to-frequency conversion methods markedly enhances efficiency. Termed 'multi-scale feature fusion', this approach not only elevates learning performance but also optimizes the deployment of computational resources.

3. The Methods of Image Feature Extraction

3.1. Definition of Feature Extraction

Feature extraction refers to the process of extracting high-level features from data that are representative in certain aspects or embody certain patterns. In super-resolution tasks, the features correspond to positions where pixel values exhibit abrupt changes, such as edges and textures. If appropriate method can be deployed to attain information with high frequency, then the appreciated features can be found, and the insignificant features be ignored or reduced, the training time of neural networks will be significantly reduced.

3.2. Classification of Feature Extraction

3.2.1. Wavelet transform

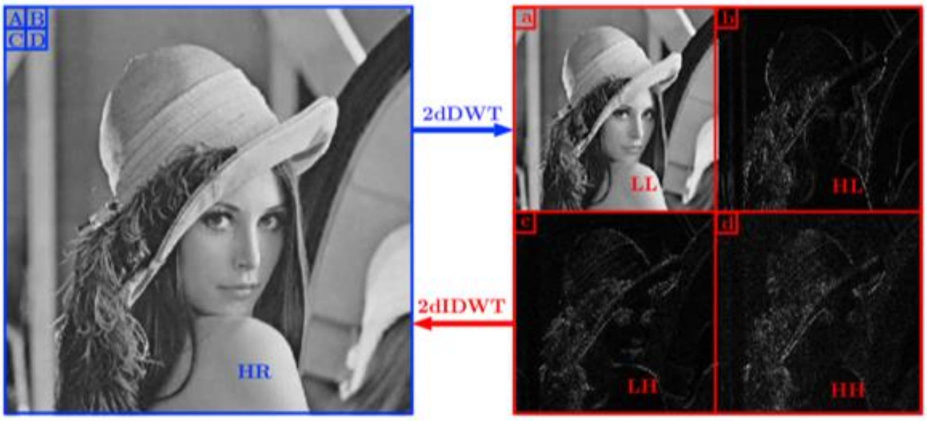

Fig.2 The wavelet-transform (Photo/Picture credit: Original).

Wavelet transform is a traditional technique for signal processing. It facilitates the transformation from the time domain to the frequency domain. As illustrated in Figure 2, through two-dimensional discrete wavelet transform (2D DWT), an image can be transformed into one low-frequency component (LL) and three high-frequency components (HL, H, H).

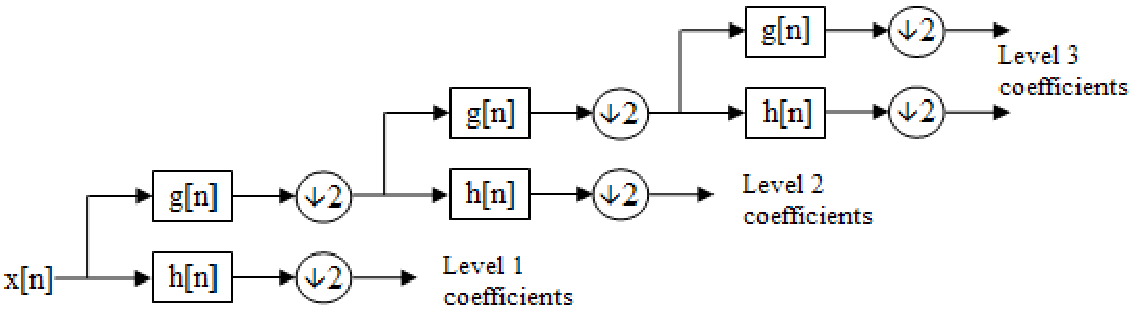

Fig.3 (Photo/Picture credit: Original).

The frequency seperation function for the process in Figure 3is given as:

$$x_{\alpha,L}[n] = \sum_{k=0}^{K-1} X_{\alpha-1,L}[2n-k]g[k]$$
$$x_{\alpha,H}[n] = \sum_{k=0}^{K-1} X_{\alpha-1,L}[2n-k]h[k] \quad (1)$$

Where the x[n] refers to Input signal or the original image, and the g[k] is defined as a low-pass filter, and h[k] is defined as a high-pass filter. The scaling factor is 2, which means after every single 2dDWT, The size of the image is reduced to half of its original size. Here, L1, L2 and L3 are defined to represent wavelet coefficients of different levels respectively. L1 coefficients contain the coarsest approximation, while L3 coefficients contain the finest detail information.

3.2.2. The Fourier transform

The Fourier transform simplifies the analysis of complex problems involving the time of signals and their corresponding spectrum in the research field. It transforms complex problems in the time domain into simpler ones in the frequency domain, thereby simplifying the analysis process. The Fourier transform encompasses various types, it includes not only continuous Fourier transform, discrete Fourier transform (DFT), but also fast Fourier transform (FFT) and short-time Fourier transform (STFT), among others. However, in the context of super-resolution tasks, the discrete Fourier transform (DFT) is commonly employed [16,17].

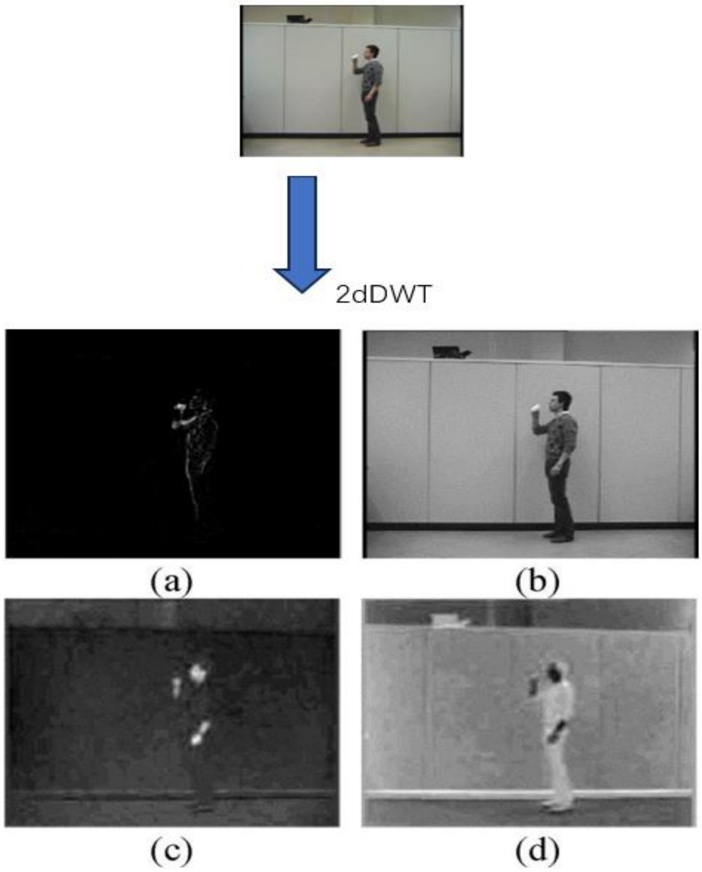

Fig. 4 (Photo/Picture credit: Original).

The 2dDFT function for the process in Figure 3 is given as:

$$
\begin{aligned}
&F(u,v)=\frac{1}{\sqrt{MN}}\sum_{x=0}^{M-1}\sum_{y=0}^{N-1} f(x,y)\exp\left[-\mathrm{j}2\pi\left(\frac{xu}{M}+\frac{yv}{N}\right)\right]\\
&u=0,1,\cdots,M-1;v=0,1,\cdots,N-1\\
&f(x,y)=\frac{1}{\sqrt{MN}}\sum_{u=0}^{M-1}\sum_{v=0}^{N-1} F(u,v)\exp\left[\mathrm{j}2\pi\left(\frac{xu}{M}+\frac{yv}{N}\right)\right]\\
&x=0,1,\cdots,M-1;y=0,1,\cdots,N-1
\end{aligned}
\tag{2}
$$

The topmost image represents the original image. Below are images depicting four different features. Among them, (a) represents the dynamic feature of the original image: motion feature; (b) displays the brightness characteristics of the original image; (c) and (d) describe two opposing color features. The 2D DFT formula in Figure 4 illustrates the transformation process. Where the F (ų, v) represents the function obtained after Fourier transformation. It signifies the signal or function in the frequency domain. The M, N denotes the dimensions of the image or signal, i.e., the number of columns and rows in the image. The f (x, y) is the original signal or image, typically represented in the spatial domain. It is the object of Fourier transformation.

4. The Methods of Image Feature Fusion

4.1. Definition of Feature Fusion

Shallow layers in Convolutional Neural Networks (CNNs) specialize in capturing low-level features like colors and edges, while deeper layers are adept at learning high-level representations, such as the properties of objects. Combining the insights from both shallow and deeper layers, integrating low-level details with high-level semantics, becomes imperative for achieving successful High-Resolution (HR) reconstruction tasks.

4.2. Multi-scale Image Feature Fusion

Most models are designed as single-channel networks, which limits their performance in handling high-frequency information in input images. When input images contain less high-frequency information, these models often fail to accurately predict high-resolution images. This is because their network structures can only learn training images with high-frequency information along one pathway, and cannot effectively capture and utilize other detailed information in the input images, Dargahia et al proposed a three-path network in 2021, which has the same subnetworks fed by different Low Resolution (LR) images. This approach utilizes multi pathway to handle both the original low-resolution (LR) data and its high-frequency discarded version, thereby encouraging the network to learn advanced features from low-information inputs. Additionally, employing a multi-scale approach allows for the simultaneous detection of coarse and fine features, thus enhancing the network's performance. Through parallel feature extractors, residual learning, and multi-scale feature reconstruction, the network can effectively acquire short-term and long-term feature maps, leading to improved predictions of high-resolution images [18].

4.3. AMFFN

In recent years, with the continuous advancement of technology, the issue of image degradation has become increasingly complex, posing new challenges to the processing of remote sensing images, especially in inferring high-frequency details. To address this challenge, Wang et al proposd an adaptive multi-scale feature fusion network (AMFFN) for remote sensing image super-resolution [19]. It is a self-adaptive multi-scale feature fusion network specifically designed for handling super-resolution of remote sensing images. AMFFN leverages multiple AMFE modules, squeeze-excitation mechanisms, and adaptive gating mechanisms to extract and fuse feature information from low-resolution images. Subsequently, these feature details are successfully utilized for reconstructing high-resolution images using sub-pixel convolutional methods. The integration of squeeze-excitation mechanisms and adaptive gating mechanisms aids in learning channel correlations of feature maps and intelligently determining the extent to retain previous feature information, thereby reducing redundancy within intermediate multi-scale features and enhancing the utilization efficiency of useful feature information. This innovative approach provides a fresh perspective and toolset for addressing key challenges in remote sensing image processing.

4.4. Combinations of AM and FET

The combination of attention mechanisms (AM) and feature extracting techniques (FET) can be also regarded as multi-scale fusion method. The convolution network (CNN) often introduces convolution operation to extract the feature map from LR pictures, Aycy et al proposed a section of feature extracting using ConvNet, which performs the finetuning of ConvNet as a fixed feature extractor, as it is a pretrained model, however, debugging parameters for this task can be very slow because, in this architecture, the feature extractor model needs to be pretrained, and training takes time. However, this is unnecessary because many method such as time-frequency-domain transformation methods already have excellent feature extraction properties, such as wavelet transformation [20]. It can extract features from four dimensions of the image separately, and these four features can also undergo wavelet transformation again. Therefore, researchers can decide the number of decomposition layers according to the needs of their own tasks, which can greatly enhance the versatility of the model and reduce computational burden. As for a multiscale aspect, feature extraction is usually combined with attention mechanisms, which allows the system to focus attention on learning specific information from the feature map. If a residual module is added, it can greatly reduce computational burden. This type of multiscale fusion technique can significantly reduce model parameters, computational burden, and achieve state-of-the-art performance [21].

4.5. DWTrans

RefSR technology enhances the content and texture information of LR images using HR reference images, but finding high-quality reference images is time-consuming and laborious. Li et al. proposed the self-referenced image super-resolution (DWTrans) method, which utilizes pre-trained diffusion big models and window-adjustable transformers to offer a new solution to this problem.

In RefSR methods, the key to network performance lies in effectively transferring valuable features from reference to LR images. Current approaches employ a matching-fusion strategy, but performance deteriorates when dealing with differences in brightness, contrast, and color distribution. To address this, a novel fusion strategy is proposed, accompanied by the introduction of an Adaptive Deformable Fusion Module (ADFM) in WAT. This module aids in enhancing network performance, ensuring better integration of useful features from SRef images into LR images, thereby improving the quality and accuracy of image super-resolution.

4.6. RGT

Transformer topologies have shown promise in terms of superresolution (SR) of images. Because of Transformer's selfattention's (SA) quadratic computational complexity, most current approaches use SA locally to cut costs. On the other hand, the global context exploitation—which is essential for precise picture reconstruction—is limited by the local design. The Recursive Generalization Transformer (RGT) for image SR was proposed by Chen et al. in this study. It is appropriate for high-resolution pictures and has the ability to collect global spatial information. Chen and colleagues introduced the recursive-generalization self-attention (RG-SA) paradigm. After creating representative feature maps by recursively aggregating input features, cross-attention is used to extract global information [22].

According to He et al. (2016), the logical concept involves utilizing the vanilla skip connection to directly integrate the input and output properties. Suggesting hybrid adaptive integration (HAI) as a solution to address the aforementioned challenges.

5. Attention Mechanism

5.1. Theoretical Background of Attention Mechanisms

In the nascent stages of the invention of the attention mechanism, its introduction aimed to emulate the attention mechanisms within the human visual system, enabling machines to 'focus attention' on specific portions of input data. This 'specific portion,' often holding significant weight, pertains to either the presentation quality or the influence of image details. Within the realm of deep learning, such attention mechanisms enable neural networks to prioritize crucial elements over others, thereby optimizing the efficiency of the network's learning process and substantially reducing the parameter requirements within the neural network architecture.

5.2. Types and Classifications

To advance the application of attention mechanisms in deep learning, numerous methods based on attention mechanisms have been proposed. However, mainstream academic perspectives categorize attention mechanisms into two primary types: channel attention mechanisms and non-local attention. Channel Attention. To emphasize crucial information among channels and suppress irrelevant information, Liu et al. recently proposed a method utilizing channel attention: initially extracting global information from feature maps through global average pooling, followed by employing twain fully connected layers to study the relationships among channels [23]. Consequently, the interconnected channel-level information is obtained, thereby significantly enhancing efficiency during training. The channel attention mechanism has been the subject of extensive discussion for a considerable duration, with one of the most representative and seminal approaches being the quite deep residual channel attention network (RCAN) introduced by Zhang et al Addressing the challenge of treating a plethora of insignificant information equally with crucial high-frequency information, RCAN acknowledged the interdependency among different channels [24]. By incorporating the residual in residual (RIR) module, RCAN effectively disregards low-frequency information within the feature maps, enabling the network to focus solely on learning information with high-frequency, thereby enhancing training efficiency.

Non-local Attention. Its mechanism was initially devised to address the challenge of modeling long-range dependencies. Traditional attention mechanisms, such as channel attention, typically consider only the correlations between locally adjacent positions, thus overlooking relationships between distant positions. However, in numerous computer vision and natural language processing tasks, dependencies between distant positions are often crucial, particularly when handling long-term dependencies or global contextual information. Building upon this theoretical foundation, numerous methods utilizing non-local attention for addressing super-resolution problems have been proposed [25]. However, these approaches often suffer from either limited coverage of the non-local region or immense computational demands. In response to this challenge, Mei et al. introduced a novel Non-Local Sparse Attention (NLSA) mechanism, significantly enhancing information propagation across dimensions while markedly reducing computational overhead [26]. Recently, Li et al. further proposed a design that utilize up to four types of patch crop modes called shift-window-based nonlocal attention block, in each block, two cascaded convolution blocks (i.e., consist of the convolution layer, batch normalization, LeakyReLU activation function) are adopted, as a result the optimization process is not only stabilized but also results in a reduction of model parameters [27].

5.3. Enhancing Feature Extraction with Attention

To accomplish the agenda of attention mechanisms for enhancing feature extraction, attention mechanism are often integrated with other methods. For instance, Shan et al. integrated convolutional neural networks (CNNs), residual learning networks, and attention mechanisms in their research. The utilization of residual learning networks significantly reduces the sparsity of feature maps, thereby positively enhanced training speed [28]. When combined with attention mechanisms, further parameter reduction can be achieved, allowing the neural network to prioritize information deemed important. In Shan's study, a Residual Attention Module was proposed as an addition of this integration, which can be also called multi-scale fusion. The attention unit is accessed via skip links, with the incorporation of partial skip connections to disseminate input information across the mapping process. This integration effectively addresses the challenge of feature disappearance induced by the inclusion of the attention mechanisms, and thereby increase the generality of attention mechanisms.

5.4. Integrating Attention in Feature Fusion

Feature fusion involves integrating features from different levels or sources, and there are various fusion methods available. Some features exert a greater influence on network learning than others, and attention mechanisms are great at discerning the importance of feature maps. Therefore, attention mechanisms play a crucial role in feature fusion by effectively discerning the importance of different features.

6. Conclusion

Empirical evidence robustly supports the assertion that feature optimization methodologies markedly diminish the computational load during the training phase. The synergistic integration of time-frequency transformation techniques and attention mechanisms considerably amplifies their effectiveness, thereby streamlining the model and significantly enhancing training efficiency. Specifically, the melding of wavelet transformation with the intrinsic sparsity of feature maps, when augmented through an attentional channel mechanism, yields a more computationally efficient process, effectively lessening the computational burden. These methods exhibit a versatile applicability across a wide array of deep learning network architectures. For instance, the amalgamation of residual networks with these feature optimization strategies, or with alternative multi-scale fusion techniques, facilitates an adept learning interpolation between inputs and outputs, predicated on pre-existing data sparsity. Given the inherently sparse nature of the input data, the integration of residual modules further intensifies this sparsity, thereby abbreviating the training duration. The considerable reduction in computational demands attributable to feature optimization offers substantial advantages. It paves the way for deploying neural networks on embedded devices without necessitating advanced hardware. This is particularly beneficial in the context of the burgeoning 5G internet technology, which has significantly elevated the data throughput capabilities of mobile devices.

Authors Contribution

All the authors contributed equally and their names were listed in alphabetical order.

References

[1] Umirzakova, S., Ahmad, S., Khan, L. U., & Whangbo, T. (2024). Medical image super-resolution for smart healthcare applications: A comprehensive survey. Information Fusion, 103.

[2] (2023). Medical image super-resolution reconstruction algorithms based on deep learning: A survey. Computer Methods and Programs in Biomedicine, 238.

[3] Last, M. G. F., Voortman, L. M., & Sharp, T. H. (2023). scNodes: A correlation and processing toolkit for super-resolution fluorescence and electron microscopy. Nature Methods, 20, 1445-1446.

[4] (2023). Lightweight image super-resolution based on deep learning: State-of-the-art and future directions. Information Fusion, 94, 284-310.

[5] Wang, T., Sun, W., Qi, H., & Ren, P. (2018). Aerial image super-resolution via wavelet multiscale convolutional neural networks. IEEE Geoscience and Remote Sensing Letters, 15(5), 769-773.

[6] Wang, Z., Zhao, Y., & Chen, J. (2023). Multi-Scale Fast Fourier Transform Based Attention Network for Remote-Sensing Image Super-Resolution. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 16, 2728-2740.

[7] Liu, B., Zhao, L., Liu, W., & Li, Y. (2024). MWLN: Multilevel Wavelet Learning Network for Continuous-Scale Remote-Sensing Image Super-Resolution. IEEE Geoscience and Remote Sensing Letters, 21, 1-5.

[8] Chen, H., Huang, L., Liu, T., & Ozcan, A. (2023). eFIN: Enhanced Fourier Imager Network for Generalizable Autofocusing and Pixel Super-Resolution in Holographic Imaging. IEEE Journal of Selected Topics in Quantum Electronics, 29(4), 1-10.

[9] Chen, Y., Liu, L., Phonevilay, V., et al. (2021). Image super-resolution reconstruction based on feature map attention mechanism. Applied Intelligence, 51, 4367–4380.

[10] Liu, B., & Chen, J. (2021). A Super Resolution Algorithm Based on Attention Mechanism and SRGANNetwork. IEEE Access, 9, 139138-139145.

[11] Yang, X., Li, X., Li, Z., & Zhou, D. (2021). Image super-resolution based on deep neural network of multiple attention mechanism. Journal of Visual Communication and Image Representation, 75.

[12] Liu, Y., Yang, D., Zhang, F., et al. (2023). Deep recurrent residual channel attention network for single image super-resolution. Visual Computer.

[13] Zhang, Y., Li, K., Li, K., Wang, L., Zhong, B., & Fu, Y. (2018). Proceedings of the European Conference on Computer Vision (ECCV), 286-301.

[14] Hsu, W. -Y., & Jian, P. -W. (2023). Wavelet Pyramid Recurrent Structure-Preserving Attention Network for Single Image Super-Resolution. IEEE Transactions on Neural Networks and Learning Systems.

[15] Dai, T., Cai, J., Zhang, Y., Xia, S.-T., & Zhang, L. (2019). Second-order attention network for single image super-resolution. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 11065–11074.

[16] Zhang, Y., Li, K., Li, K., Zhong, B., & Fu, Y. (2019). Residual non-local attention networks for image restoration. arXiv preprint arXiv:1903.10082.

[17] Mei, Y., Fan, Y., Zhou, Y., Huang, L., Huang, T. S., & Shi, H. (2020). Image super-resolution with cross-scale non-local attention and exhaustive self-exemplars mining. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 5690–5699.

[18] Mei, Y., Fan, Y., & Zhou, Y. (2021). Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 3517-3526.

[19] Li, S., Tian, Y., Wang, C., Wu, H., & Zheng, S. (2023). Hyperspectral image super-resolution network based on cross-scale nonlocal attention. IEEE Transactions on Geoscience and Remote Sensing, 61, 1-15.

[20] Shan, L., Bai, X., Liu, C., et al. (2022). Super-resolution reconstruction of digital rock CT images based on residual attention mechanism. Advances in Geo-Energy Research, 6(2), 157-168. https://doi.org/10.6690/ager.2022.02.07

[21] Chen, Y., Xia, R., Yang, K., et al. (2024). MFFN: Image super-resolution via multi-level features fusion network. Visual Computer, 40, 489–504. https://doi.org/10.1007/s00371-023-02795-0

[22] An, T., Zhang, X., Huo, C., Xue, B., Wang, L., & Pan, C. (2022). TR-MISR: Multiimage super-resolution based on feature fusion with transformers. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 15, 1373-1388.

[23] Niu, A., et al. (2022). MS2Net: Multi-scale and multi-stage feature fusion for blurred image super-resolution. IEEE Transactions on Circuits and Systems for Video Technology, 32(8), 5137-5150.

[24] Dargahi, S., Aghagolzadeh, A., & Ezoji, M. (2021). Single image super resolution using multi-path convolutional neural network. In 2021 5th International Conference on Pattern Recognition and Image Analysis (IPRIA) (pp. 1-6). Kashan, Iran.

[25] Wang, X., Wu, Y., Ming, Y., & Lv, H. (2020). Remote sensing imagery super resolution based on adaptive multi-scale feature fusion network. Sensors (Basel, Switzerland), 20.

[26] Avci, D., Sert, E., Dogantekin, E., Yildirim, O., Tadeusiewicz, R., & Plawiak, P. (2023). A new super resolution Faster R-CNN model based detection and classification of urine sediments. Biocybernetics and Biomedical Engineering, 43(1), 58-68.

[27] Li, G., Xing, W., Zhao, L., Lan, Z., Sun, J., Zhang, Z., Zhang, Q., Lin, H., & Lin, Z. (2023). Self-reference image super-resolution via pre-trained diffusion large model and window adjustable transformer. In Proceedings of the 31st ACM International Conference on Multimedia (MM '23) (pp. 7981–7992). Association for Computing Machinery.

[28] Chen, Z., Zhang, Y., Gu, J., Kong, L., & Yang, X. (2023). Recursive generalization transformer for image super-resolution. ArXiv. abs/2303.06373.