

Existing Approaches for Single-Image Super-Resolution

Subjects: Computer Science, Artificial Intelligence

Contributor: Khanh Hung Vu, Duc Phuc Nguyen, Duc Dung Nguyen, Hoang-Anh Pham

Deep learning has been introduced to single-image super-resolution (SISR). These techniques have taken over the benchmarks of SISR tasks. Nevertheless, most architectural designs necessitate substantial computational resources, leading to a prolonged inference time on embedded systems or rendering them infeasible for deployment.

Keywords: single-image super-resolution (SISR) ; deep learning ; quantization

1. CNN-Based Methods

The most primitive and pioneering method using CNN is the SRCNN, proposed by Dong et al. ^[1], which has proven to be superior to traditional non-deep learning methods in terms of reconstructed image quality. This study also shows that the normal sparse-coding-based image recovery model can be viewed as a depth model. However, the three-layer network is unsuitable for recovering compressed images, especially when dealing with blocking artifacts and smooth regions. When different artifacts are concatenated together, the features extracted by the first layer are noisy, which causes unexpected noise patterns during the reconstruction.

Because of the real-time advantage of the SRCNN model, and to overcome the computational disadvantage, Ahn et al. ^[2] proposed a model that can be implemented directly on the FPGA, named Optimal-FSRCNN. This model has used the Transforming DeConvolutional (TDC) layer within the Convolutional Layer method to convert the deconvolution layer to the equivalent convolution layer, to overcome the inherent overlapping sum problem, which causes increased latency, consumes a lot of power and other hardware resources, and maximizes real-time super-resolution image parallelization using lightweight deep learning.

Inheriting the improvement of the above disadvantages, another model was formed to contribute to their improvement, which is the LAPSRN model proposed by Lai et al. ^[3]. Deep supervision with the Charbonnier loss function improves performance through better handling of outliers. Therefore, the model has a great ability to learn complex mappings and is effective in minimizing undesirable visualizations. Furthermore, learning upsampling filters not only minimize the generation of reconstruction artifacts produced by bicubic interpolation but also contribute to minimizing computational complexity. Experimental results show that this model is capable of solving the story of time. However, the model size is still relatively large. To reduce the number of parameters, one can replace the deep convolution layers at each level with recursive layers. In terms of image quality, not only LAPSRN, but also most other parametric SR methods fail to recover fine structure.

2. Distillation Methods

Although the CNN-based models achieve outstanding performance, the proposed networks still have disadvantages. To achieve better performance, one needs to design a deeper network. However, as a result, these methods are computationally expensive and consume large amounts of hardware resources, which are rarely applied in mobile and embedded applications.

To solve that problem, some researchers have proposed some models to meet the needs. First, Hui et al. ^[4] proposed the IDN model, which extracts more helpful information with fewer convolutional layers. Although IDN has reduced parameters compared to the previous method, this reduction is achieved at the cost of significant performance sacrifice. Then, they proposed an alternative model, IDMN ^[5], based on their previous work IDN, with a more lightweight structure and faster running. The IMDN model uses an information multi-distillation block (IMDB) to further improve performance in both PSNR and inference time; this model took first place in the AIM 2019 constrained image super-resolution challenge

[6]. However, the number of parameters in the IMDN model is greater than most lightweight SR models, such as VDSR [7] and IDN. However, the IMDN model still has room for improvement to be more lightweight.

The main component of both IDN and IMDN is an *Information Distillation Mechanism* (IDM) that explicitly divides previously extracted features into two parts: retained and refined. On the other hand, IDM is not efficient enough and brings inflexibility to network design. It is not easy to combine identity connections with IDM. However, models using this approach can be geared towards real-time systems because such models are often very flexible in making the trade-off between PSNR and inference time, via a parameter called *Channel Modulation Coefficient*.

3. Attention-Based Methods

The authors in [8] introduced SENet, which uses channel attention to exploit interdependencies between channels of a model, improving feature map efficiency. This CNN-based squeeze-and-excitation network enhances classification networks and is now widely used in neural network design for down-streaming computer vision tasks [9].

Channel attention mechanisms have been introduced to improve the performance of neural networks in the image super-resolution domain. Zhang et al. [10] developed a CNN model called RCAN that utilizes channel attention to address SISR problems. RCAN combines residual in residual (RIR) and channel attention (CA), in which RIR is used to propagate low-frequency information [11] from the input to the output, allowing the network to learn residual information at a coarse level. The deep architecture of RCAN, with over 400 layers, enables the network to learn deeply and achieve high performance.

Super-resolution algorithms aim to restore mid-level and high-level frequencies because the low-level frequencies can be obtained from the input LR image without the need for highly complex computations. The RCAN model models the features equally or on a limited scale, ignoring the abundantly rich frequency representation at other scales. As such, these lack discriminative learning capability and capacity across channels, limiting convolutional neural network capabilities. To overcome this limitation, Anwar et al. [12] proposed the DRLN network, which uses dense connections between RBs to utilize previously computed features. The model also uses the Laplacian pyramid attention to weigh the features at multiple scales and according to their importance. The DRLN network has fewer convolutional layers than the RCAN model but more parameters. Nevertheless, it is computationally efficient due to the multiplexing of the channels, contrarily to RCAN, which uses a more expensive operation that involves more channels.

Channel attention-based approaches in image super-resolution have limitations in preserving texture and natural details due to the processing of feature maps at different layers, which can result in the loss of details in the reconstructed image. To address this issue, Niu et al. [13] proposed the HAN network, which can discover correlations between hierarchical layers, channels within each layer, and the positions of each channel, thereby activating the representative power of CNN. The HAN model also proposes a LAM model to demonstrate the relationship between features at hierarchical levels to promote CNN's performance. Additionally, the CSAM module improves the discriminative learning of the network. However, LAM only assigns a single importance weight to all features in the same class and does not consider the difference in the spatial positions of these features.

4. Feedback Network-Based Methods

The feedback mechanism differs from conventional input-to-target mapping, incorporating a self-correcting phase during the model's learning process. In computer vision, feedback mechanisms have become increasingly popular in recent years. In particular, feedback mechanisms are commonly used in SR models because they can transfer in-depth information to the front end of the network to help process shallow information more effectively. This aids in the reconstruction of HR images from LR images.

Haris et al. [14] proposed a method called DBPN to capture the interdependencies between LR and HR image pairs. The method uses iterative back projection to calculate reconstruction error and extract high-frequency features, which are then merged to improve the accuracy of the HR image. DBPN alternates between the upsampling and downsampling layers and improves the performance through dense connections, especially for magnification, which improves by a factor of eight. However, this method is computationally expensive and increases network complexity and inference time.

Zhen et al. [15] developed the SRFBN network, which employs the negative feedback mechanism of human vision to enhance low-level representations with high-level information. The intermediate states in a constrained RNN are used to implement this feedback mode. The feedback blocks are designed to handle the feedback wiring system and generate high-level information more efficiently. The SRFBN network improves the reconstruction performance using few

parameters to reduce the likelihood of overfitting due to the feedback mechanism. Still, this approach leads to an increase in computational costs. However, networks such as DBPN and SRFBN cannot learn feature mapping at multiple context scales.

The feedback mechanism used in SRFBN only propagates the highest-level feature to a shallow layer, leaving out other high-level information captured in different receptive fields. As a result, SRFBN does not make full use of high-level features or adequately refine low-level features. To address these drawbacks, Li et al. [16] introduced the GMFN network, which transfers refined features to the shallow layers. This model assumes that enough contextual information can refine the basic layers. The feature maps extracted at different layers contain complementary information for image reconstruction and are captured in different receptive fields. Then, the feedback connection optimizes the basic information with the help of the advanced counterpart.

Liu et al. [17] developed the HBPN network, taking inspiration from the DBPN model. It employs residual hourglass modules in a hierarchical structure to improve error estimation and achieve superior results. However, the ability of these models to generalize is restricted by the kernel set k and scaling factors s . Furthermore, the HBPN model requires both RGB and YUV images as input, resulting in considerable computational overhead.

The authors in [18] also introduced an ABPN network that follows the concept of the HBPN model and utilizes RBPB blocks to expand the receptive field of back-projection blocks. By leveraging the original information in the LR input, RBPB can enhance the SR performance by exploiting the interdependencies between the LR input and the SR output. However, ABPN has some limitations. Firstly, it fails to merge high-frequency features. Secondly, the standard convolutional layers and self-attention modules do not distinguish between different degrees of feedback errors, resulting in back-projection blocks being unable to focus on areas with significant errors and reducing the correction effect.

5. Recursive Learning-Based Methods

The feedback-based model utilizes self-correcting parameters, distinguishing it from the recursive learning-based model, where the parameters are shared among the modules.

The CARN network proposed by Ahn et al. [19] uses a lightweight model that replaces the standard RB block with an efficient RB version, which has fewer parameters and computational costs than RB but similar learning capabilities. The CARN network achieved super-resolution benchmark results among lightweight models with a parameter count of less than 1.5 million. Still, the performance is limited by the number of parameters, and the PSNR and SSIM metrics are reduced.

To reduce computational complexity and cost, Choi et al. [20] proposed the BSRN network, which includes an initial feature extractor, a recursive RB, and an upscaling part. The SRRFN model, on the other hand, was proposed by Li et al. [21] and achieved superior results with fewer parameters and less execution time than the RCAN model. SRRFN introduced a new fractal module (FM), which can create multiple topological structures based on a simple component to detect rich image features and increase the fault tolerance of the model.

The LP-KPN network proposed by Cai et al. [22] is based on the Laplacian pyramid to learn kernels per pixel for the decomposed image pyramid, to achieve high computational efficiency with large kernel sizes. The LP-KPN model outperforms the CRAN models trained on simulation data while having fewer convolutional layers. However, the LP-KPN model only reconstructs the LR image by collaborating with different pixel-local reconstructions, which does not fully use hierarchical features across different frequencies.

6. GAN-Based Methods

GANs [23] uses a game theory approach that includes the generator and the discriminator trying to fool each other. The generator generates SR images that the discriminator cannot distinguish as real or artificial HR images. In this way, HR images with better perceptual quality are produced. The corresponding PSNR values are often attenuated, highlighting the problem that the quantitative measures common in the SR literature do not encapsulate the perceptual accuracy of the generated HR outputs.

GAN models overcome the weakness when using the loss function MSE as the criterion. Although minimizing MSE also maximizes PSNR and is a common metric used to evaluate and compare SR algorithms, the ability of MSE (and PSNR) to capture relevant differences in terms of perceptions, such as high texture details, is very limited because they are determined based on differences in the image pixels. The higher PSNR does not necessarily reflect a better perceptual

outcome. Realizing the above, Ledig et al. [24] proposed the generative model related to GAN in the super-resolved problem as SRGAN. The SRGAN model uses a deep residual network (ResNet) with skip connections and diverges from the MSE as the sole optimization objective. The model also identifies a new perceptual loss function using VGG network high-level feature mappings combined with a discriminator that encourages solutions that are difficult to distinguish from HR images. Although SRGAN significantly improves the overall image quality of the reconstruction, its disadvantage is that the model is difficult to train, often producing artifacts in SR images.

To avoid the generation of SR images with artifacts, Xintao Wang et al. [25] proposed the ESRGAN network to make the reconstructed image more realistic. Firstly, this model removes the BN layer, reducing computational costs and memory. Moreover, it also contributes to reducing the artifacts in the SR image. Secondly, the use of pre-activation features results in a more accurate brightness distribution (i.e., closer to the actual brightness), producing sharper edges and richer textures. Thirdly, the deep learning model shows outstanding performance in easy training thanks to the RRDB block without the BN layer. Even so, the model has difficulty recreating the high-frequency edges. Furthermore, the regeneration effect greatly deteriorates when the ESRGAN model is applied to multiple degradations. One form of the SRGAN model with a slight variation that gives the most satisfactory results in terms of inference time and applies to low-end devices, such as embedded systems, is the SwiftSRGAN model, proposed by Koushik Sivarama Krishnan et al. [26], so that the model can run on a time-constrained system; this model changes the convolution block to a DSC block.

The generator structure of the SRGAN model uses a deep residual network called SRResNet. This type of architecture gives good results in terms of structural similarity and detail. However, in experiments, it was found to perform poorly in maintaining global information and high-level structure, sometimes distorting the overall characteristics of the image. Therefore, Mirchandani et al. [27] proposed the DPRSGAN model. To avoid the above problem, the model uses dilated convolution to capture global structures and fewer parameters. On the other hand, changing the discriminator to a Markovian discriminator (PatchGAN) speeds up model training and produces sharper details.

Additionally, in order to deal with the degraded images in the real world in general, Real-ESRGAN was formed, as proposed by Wang et al. [28]. Compared to ESRGAN models, Real-ESRGAN is trained entirely on synthetic data, which helps it to recover complex real-world images with better image performance. Another model used for the real-world image super-resolution problem is BSRGAN, proposed by Zhang et al. [29]. It has been proposed as a real-world degradation model to remove the disadvantages of synthetic data generation and build a robust model for different combinations of downsampling kernels, blur kernels, and noise. The method demonstrates outstanding results when dealing with real-world datasets where the degradation model is unknown. Nevertheless, since the models are trained using pairs of images generated by such real-world degradation models and considered for general scenarios, it can be confirmed that the denoising has surpassed the required level. Hence, the real-world degradation model is not appropriate for specific visual inspection tasks that require fixed HR and LR noise levels for noise reduction.

In addition, F2SRGAN [30] enhances the receptive field of the convolution operator by employing Fast Fourier Convolution. This technique enables the model to capture low-frequency characteristics in the frequency domain, resulting in quicker coverage of high-frequency features compared to the traditional spatial domain of standard convolution.

7. Transformer-Based Methods

Methods employing CNN or GAN are limited to using local image information, ignoring the global interaction between image components, resulting in low-quality recovery. Transformer is a novel deep learning model that employs a self-attention mechanism based on assigning varying weights to the significance of each portion of the input data. It has been a cutting-edge method in natural language processing (NLP) since its inception. Transformer has become increasingly prominent in computer vision tasks such as object detection, segmentation, classification, and image super-resolution due to its ability to solve long-term dependency issues. It can utilize local and global information from the input image to produce a more detailed output image. This innovation has attracted many researchers and introduced a new network architecture for image super-resolution.

A typical model for applying Transformer to this super-resolution problem is the ESRT proposed by Zhisheng et al. [31]. The model consists of an LCB block, which uses HPB blocks to automatically adjust the feature map size to extract intensive feature mappings with low computational cost. The model also has an LTB block to capture long-term dependencies between similar patches in an image with the help of specially designed Efficient Transformer (ET) and Efficient Multi-Head Attention (EMHA) mechanisms. This model is proposed to effectively enhance the feature representation and long-term dependence of similar patches in an image, to achieve better performance with low

computational cost. Although Transformer is a powerful model, there is still the problem that Transformer-based models are heavy models, i.e., the number of parameters and the amount of data to train are still large.

8. Frequency-Domain Based Methods

Super-resolution (SR) algorithms that use a frequency domain-based method transform low-resolution (LR) input images into the frequency domain to estimate an HR image. The reconstructed high-resolution (HR) image is then transformed back into the spatial domain. Fourier and wavelet transform-based methods are two algorithms that depend on the transformation used to convert images to the frequency domain. In [32], the authors converted LR satellite images to the frequency domain by using the discrete Fourier transform (DFT). Then, the relationship between the aliased DFT coefficients of the LR images and those of the unknown HR image was combined. In addition to enhancing the high-frequency information of images, frequency domain-based SR techniques have low computational complexity. However, these methods have some drawbacks, including being insufficient to handle real-world applications and having difficulty expressing prior knowledge used to regularize the SR problem. Many frequency domain approaches rely on Fourier transform properties such as the shifting and sampling theorems, making them easy to understand and apply. Some frequency domain methods make assumptions that enable the use of efficient procedures for computing restoration, such as the Fast Fourier Transform (FFT).

Implicit neural functions, parameterized by multilayer perceptrons (MLP), have shown great success in representing continuous domain signals such as images, shapes, and signals. However, one drawback of using a standalone MLP is that it tends to focus on low-frequency components [33] and may not capture fine details [34]. To address this limitation, Lee et al. [35] introduced a dominant frequency estimator called the LTE tool, which improves the input features of the MLP. The LTE model includes three additional trainable layers that process the encoder output and correspond to sine and cosine waves' amplitude, frequency, and phase. This output is then used as input for an MLP that has four fully connected layers. To further enhance the results, a global skip connection with a bilinear upsampled version of the input is added to the entire model, allowing the deep model to focus on the residual between the closed-form approximation and the final result. While LTE has achieved high-quality arbitrary-scale rectangular super-resolution with high-frequency details, its spatially varying nature prevents it from evaluating a frequency response for image warping.

Zhang et al. [36] developed a model called SwinFIR that combines the SwinIR model proposed by Liang et al. [37] with the FFC block [38]. This model uses a frequency domain approach to capture global information better and restore high-frequency details in images. The SwinFIR model performs well in restoring images with periodic transformations and challenging samples. However, one limitation of this model is that it is slow for large-scale images, as measuring importance at a global spatial scale requires vector multiplications along rows and columns. Recently, Nguyen et al. [30] proposed an enhanced model F2SRGAN that further improves the FFC block by performing the convolution operator directly in the frequency domain, rather than splitting the real and imaginary parts and implementing them separately.

References

1. Dong, C.; Loy, C.C.; He, K.; Tang, X. Image Super-Resolution Using Deep Convolutional Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* 2015, 38, 295–307.
2. Ahn, S.; Kang, S.J. Deep Learning-based Real-Time Super-Resolution Architecture Design. *J. Broadcast Eng.* 2021, 26, 167–174.
3. Lai, W.; Huang, J.; Ahuja, N.; Yang, M. Deep Laplacian Pyramid Networks for Fast and Accurate Super-Resolution. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, 21–26 July 2017.
4. Hui, Z.; Wang, X.; Gao, X. Fast and Accurate Single Image Super-Resolution via Information Distillation Network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Salt Lake City, UT, USA, 18–22 June 2018.
5. Hui, Z.; Gao, X.; Yang, Y.; Wang, X. Lightweight Image Super-Resolution with Information Multi-distillation Network. In *Proceedings of the 27th ACM International Conference on Multimedia*, Nice France, 21–25 October 2019; ACM: New York, NY, USA, 2019; pp. 2024–2032.
6. Zhang, K.; Gu, S.; Timofte, R.; Hui, Z.; Wang, X.; Gao, X.; Xiong, D.; Liu, S.; Gang, R.; Nan, N.; et al. AIM 2019 Challenge on Constrained Super-Resolution: Methods and Results. In *Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*, Seoul, Republic of Korea, 27–28 October 2019.

7. Kim, J.; Lee, J.K.; Lee, K.M. Accurate Image Super-Resolution Using Very Deep Convolutional Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 26 June–1 July 2016.
8. Hu, J.; Shen, L.; Sun, G. Squeeze-and-Excitation Networks. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–22 June 2018; pp. 7132–7141.
9. Xie, C.; Zeng, W.; Jiang, S.; Lu, X. Bidirectionally aligned sparse representation for single image super-resolution. *Multimedia Tools Appl.* 2018, 77, 7883–7907.
10. Zhang, Y.; Li, K.; Li, K.; Wang, L.; Zhong, B.; Fu, Y. Image Super-Resolution Using Very Deep Residual Channel Attention Networks. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018.
11. Yang, C.; Lu, G. Deeply Recursive Low- and High-Frequency Fusing Networks for Single Image Super-Resolution. *Sensors* 2020, 20, 7268.
12. Anwar, S.; Barnes, N. Densely Residual Laplacian Super-Resolution. *IEEE Trans. Pattern Anal. Mach. Intell.* 2019, 44, 1192–1204.
13. Niu, B.; Wen, W.; Ren, W.; Zhang, X.; Yang, L.; Wang, S.; Zhang, K.; Cao, X.; Shen, H. Single Image Super-Resolution via a Holistic Attention Network. In Proceedings of the 16th European Conference, Glasgow, UK, 23–28 August 2020.
14. Haris, M.; Shakhnarovich, G.; Ukita, N. Deep Back-Projection Networks For Super-Resolution. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–22 June 2018.
15. Li, Z.; Yang, J.; Liu, Z.; Yang, X.; Jeon, G.; Wu, W. Feedback Network for Image Super-Resolution. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Long Beach, CA, USA, 16–20 June 2019.
16. Li, Q.; Li, Z.; Lu, L.; Jeon, G.; Liu, K.; Yang, X. Gated Multiple Feedback Network for Image Super-Resolution. *arXiv* 2019.
17. Liu, Z.S.; Wang, L.W.; Li, C.T.; Siu, W.C. Hierarchical Back-Projection Network for Image Super-Resolution. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Long Beach, CA, USA, 16–20 June 2019; pp. 2041–2050.
18. Liu, Z.S.; Wang, L.W.; Li, C.T.; Siu, W.C.; Chan, Y.L. Image Super-Resolution via Attention based Back-Projection Networks. In Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW), Seoul, Republic of Korea, 27–28 October 2019.
19. Ahn, N.; Kang, B.; Sohn, K. Fast, Accurate, and, Lightweight Super-Resolution with Cascading Residual Network. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018.
20. Choi, J.; Kim, J.; Cheon, M.; Lee, J. Lightweight and Efficient Image Super-Resolution with Block State-based Recursive Network. *arXiv* 2019.
21. Li, J.; Yuan, Y.; Mei, K.; Fang, F. Lightweight and Accurate Recursive Fractal Network for Image Super-Resolution. In Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW), Seoul, Republic of Korea, 27–28 October 2019; pp. 3814–3823.
22. Cai, J.; Zeng, H.; Yong, H.; Cao, Z.; Zhang, L. Toward Real-World Single Image Super-Resolution: A New Benchmark and A New Model. *arXiv* 2019.
23. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative Adversarial Networks. *Commun. ACM* 2020, 63, 139–144.
24. Ledig, C.; Theis, L.; Huszar, F.; Caballero, J.; Cunningham, A.; Acosta, A.; Aitken, A.; Tejani, A.; Totz, J.; Wang, Z.; et al. Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; IEEE Computer Society: Piscataway, NJ, USA, 2017; pp. 105–114.
25. Wang, X.; Yu, K.; Wu, S.; Gu, J.; Liu, Y.; Dong, C.; Qiao, Y.; Loy, C.C. ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks. In Proceedings of the European Conference on Computer Vision (ECCV) Workshop, Munich, Germany, 8–14 September 2018; Springer International Publishing: Cham, Switzerland, 2019; pp. 63–79.
26. Krishnan, K.S.; Krishnan, K.S. SwiftSRGAN—Rethinking Super-Resolution for Efficient and Real-time Inference. In Proceedings of the 2021 International Conference on Intelligent Cybernetics Technology & Applications (ICICyTA), Bandung, Indonesia, 1–2 December 2021; pp. 46–51.

27. Mirchandani, K.; Chordiya, K. DPSRGAN: Dilation Patch Super-Resolution Generative Adversarial Networks. In Proceedings of the 6th International Conference for Convergence in Technology (I2CT), Maharashtra, India, 2–4 April 2021; pp. 1–7.
28. Wang, X.; Xie, L.; Dong, C.; Shan, Y. Real-ESRGAN: Training Real-World Blind Super-Resolution with Pure Synthetic Data. In Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops (ICCVW), Virtual, 11–14 October 2021; pp. 1905–1914.
29. Zhang, K.; Liang, J.; Van Gool, L.; Timofte, R. Designing a Practical Degradation Model for Deep Blind Image Super-Resolution. In Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops (ICCVW), Virtual, 11–14 October 2021.
30. Nguyen, D.P.; Vu, K.H.; Nguyen, D.D.; Pham, H.A. F2SRGAN: A Lightweight Approach Boosting Perceptual Quality in Single Image Super-Resolution via a Revised Fast Fourier Convolution. *IEEE Access* 2023, **11**, 29062–29073.
31. Lu, Z.; Liu, H.; Li, J.; Zhang, L. Efficient Transformer for Single Image Super-Resolution. *arXiv* 2021, arXiv:2108.11084.
32. Tsai, R. Multiframe Image Restoration and Registration. *Adv. Comput. Vis. Image Process.* 1984, **1**, 317–339.
33. Rahaman, N.; Baratin, A.; Arpit, D.; Draxler, F.; Lin, M.; Hamprecht, F.; Bengio, Y.; Courville, A. On the Spectral Bias of Neural Networks. In Proceedings of the 36th International Conference on Machine Learning, Long Beach, CA, USA, 10–15 June 2019; pp. 5301–5310.
34. Tancik, M.; Srinivasan, P.P.; Mildenhall, B.; Fridovich-Keil, S.; Raghavan, N.; Singhal, U.; Ramamoorthi, R.; Barron, J.T.; Ng, R. Fourier Features Let Networks Learn High Frequency Functions in Low Dimensional Domains. *Adv. Neural Inf. Process. Syst.* 2020, **33**, 7537–7547.
35. Lee, J.; Jin, K.H. Local Texture Estimator for Implicit Representation Function. In Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 19–20 June 2022; pp. 1919–1928.
36. Zhang, D.; Huang, F.; Liu, S.; Wang, X.; Jin, Z. SwinFIR: Revisiting the SwinIR with Fast Fourier Convolution and Improved Training for Image Super-Resolution. *arXiv* 2022, arXiv:2208.11247.
37. Liang, J.; Cao, J.; Sun, G.; Zhang, K.; Van Gool, L.; Timofte, R. SwinIR: Image Restoration Using Swin Transformer. In Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW), Virtual, 11–14 October 2021; pp. 1833–1844.
38. Chi, L.; Jiang, B.; Mu, Y. Fast Fourier Convolution. In *Advances in Neural Information Processing Systems*; Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M., Lin, H., Eds.; Curran Associates, Inc.: Red Hook, NY, USA, 2020; Volume 33, pp. 4479–4488.