Super Resolution of Images Using Residual Network

Shreyas Y J, Mallikarjun Honnalli, Sinchana S
Uday Kulkarni and Shashank Hegde
School of Computer Science and Engineering, KLE Technological University, Hubli, Karnataka, India

Keywords: SISR, Residual Networks (ResNet), Convolutional Neural Networks (CNNs), Image Processing, PSNR, SSIM.

Abstract: Single Image Super Resolution (SISR) is a vital task in computer vision that reconstructs High-Resolution (HR) images from Low-Resolution (LR) inputs. It is widely used in fields like diagnostic imaging, geospatial imaging, and video streaming. In this study, we introduce a Residual Network (ResNet) approach for super-resolution, which addresses challenges like vanishing gradients and captures finer details through deeper architectures. Our ResNet model effectively reduces computational overhead while preserving critical features, ensuring scalability across various image datasets. We evaluated its performance on the DIV2K dataset using Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM), achieving a PSNR of 30.25 dB and SSIM of 0.77. These results demonstrate that our model outperforms traditional methods and competing architectures, making it a robust solution for applications requiring high precision, such as video enhancement and real-time imaging.

1 INTRODUCTION

Single-Image Super-Resolution (SISR)(Yang et al., 2014) is a crucial task in computer vision that aims to reconstruct High-Resolution (HR) images from their Low-Resolution (LR) counterparts. It is utilized across different fields, including healthcare diagnostics, space-based observation, and video streaming.The demand for high-quality super-resolution methods has grown significantly. Traditional techniques based on interpolation, such as bicubic interpolation, often fail to preserve fine details and realistic textures, leading to visually unsatisfactory results.

The advent of deep learning, particularly Convolutional Neural Networks (CNNs)(Aloysius and Geetha, 2017), has revolutionized the field of SISR(Ye et al., 2023). Among these, Residual Networks (ResNets)(Zhang et al., 2017) has emerged as a promising architecture due to their ability to effectively mitigate issues such as vanishing gradients in deep networks while preserving critical feature information through skip connections. However, while ResNets have shown success in advanced visual tasks such as classifying image and detecting object, their direct application to low-level tasks such as SISR has been suboptimal.

In this study, we proposed a ResNet-based approach that is specifically optimized for SISR tasks. We are concerned with designing an advanced ResNet architecture tailored to the restrictions imposed by existing conventional ResNet structures while dealing with SISR-related applications. Meanwhile, we develop a scalable training framework for a multi-scale mode, which may efficiently handle several upscaling factors in a single model. Therefore, it remains adaptive and efficient under various conditions. In addition, the new approach is seriously compared with existing techniques to demonstrate that it achieves superior performance for improving image quality measures in terms of the PSNR and the SSIM.

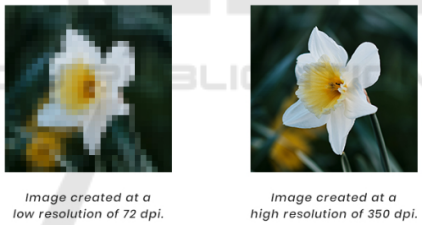

Figure 1: Difference of Low-resolution image and High-resolution image(Solutions, 2024)

The rest of this paper is structured in the following manner: Section 2 reviews related work,

Y J, S., Honnalli, M., S, S., Kulkarni, U. and Hegde, S.
Super Resolution of Images Using Residual Network.
DOI: 10.5220/0013734100004664

In *Proceedings of the 3rd International Conference on Futuristic Technology (INCOFT 2025) - Volume 3*, pages 864-870
ISBN: 978-989-758-763-4

including traditional and deep learning-based SISR approaches. Section 3 describes the proposed ResNet-based methodologies, including network design, training strategies, and multi-scale framework. Section 4 presents experimental results and comparisons with benchmark methods. Section 5 discusses the conclusions made after result, and concludes with future research directions.

2 BACKGROUND AND RELATED WORKS

Over the years, SISR has grown significantly, and most of the earlier methods were interpolation-based, such as bicubic(Khaledyan et al., 2020) and Lanczos(Bituin and Antonio, 2024). These methods may be computationally efficient but fall short of producing fine details and realistic texture in the reconstruction of images consequently leading to poor image aesthetics. Learning-based methods, for example, neighbor embedding(Wang et al., 2018) and sparse(Yang et al., 2010) coding techniques, have been used to overcome such limitations by relating LR and HR image patches. These methods were successful to a degree but were hindered by their dependence on custom features and simple architectures.

The introduction of deep learning revolutionized SISR, with CNNs(Tian et al., 2021) demonstrating outstanding performance in various tasks like picture Correction and enhancement. Early CNN-based approaches, such as Super-Resolution CNN(SRCNN)(Kumar, 2020) and Fast Super-Resolution CNN(FSRCNN) (Luo et al., 2019), showed significant improvements over conventional techniques focus on understanding complete pathways from start to finish from LR to HR images. However, these networks could not model complex image textures due to their limited depth and architectural simplicity.

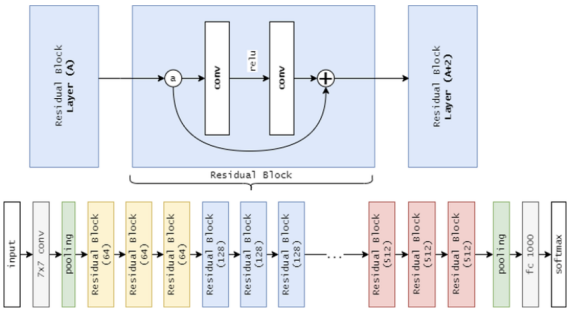

Figure 2: ResNet Architecture(Yıldırım and Dandıl, 2021)

The advent of deeper architectures like Very Deep Super-Resolution (VDSR)(Kim et al., 2016)and Super-Resolution Residual Network (SRResNet)(Ullah and Song, 2023) introduced residual learning and skip connections, enabling higher reconstruction quality while addressing issues like vanishing gradients. However, challenges continue, including scale-specific training, which causes redundancy and inefficiency, and sensitivity to hyperparameters, limiting robustness and scalability.

While residual networks (ResNets) show promise in SISR tasks, their direct application from high-level vision tasks is suboptimal. Many models include unnecessary modules like batch normalization, consuming computational resources without benefiting SISR. Furthermore, the focus on scale-specific models increases training time and memory usage, even in multi-scale approaches like VDSR(Hitawala et al., 2018).

To address these issues, we propose an optimized ResNet-based framework tailored for SISR. Our architecture improves computational efficiency and training stability by removing unnecessary components like batch normalization and incorporating residual scaling. Additionally, a multi-scale training framework with shared parameters reduces model size while maintaining performance, enhancing scalability and suitability for practical applications.

3 PROPOSED METHODOLOGY

This section provides a detailed explanation of the proposed methods, algorithms, and techniques that are used to develop the proposed ResNet-based SISR framework. It includes the architectural design, training strategies, mathematical models, implementation steps, and challenges faced during development.

3.1 Network Design

The proposed super-resolution model employs residual learning, where residual blocks learn the difference between LR and HR images. Each block has two convolutional layers with ReLU activations, with skip connections adding input to output. Omitting batch normalization improves efficiency, and scaling block outputs ($\alpha = 0.1$) stabilizes training. Stacked residual blocks enhance feature refinement, preserving essential image details for high-quality outputs.

3.1.1 Architectural Optimizations

The core feature of our architecture is residual learning, where residual blocks are used to understand the

difference between LR and HR features. Each block contains two convolutional layers followed by ReLU activations and uses skip connections to add the input directly to the output, as described in (1). To improve efficiency, batch normalization is removed, as it limits the network's ability to handle the dynamic range of image features, especially in SISR tasks. This reduction in memory usage (around 40%) allows for better reconstruction quality.

To stabilize the training of deeper networks, each residual block's output is scaled by a constant factor $\alpha = 0.1$, as formulated in (2). This scaling helps prevent large gradients during back propagation and ensures better training stability.

$$y = x + f(x, \{W_k\}), \tag{1}$$

$$y = x + \alpha \cdot f(x, \{W_k\}). \tag{2}$$

3.1.2 Final Architecture

The architecture begins with an initial convolution layer that extracts details from the low-resolution input image I_{LR}. These details are then refined through N residual blocks, where each block adjusts the features as expressed in (3). After processing through the residual blocks, upsampling is performed through pixel shuffle layers, followed by the reconstruction of the high-resolution image I_{SR} through a final convolutional layer. This approach ensures that the low-resolution input is transformed into a high-quality super-resolved image.

$$F_i = F_{i-1} + \alpha \cdot f(F_{i-1}). \tag{3}$$

After processing through the residual blocks, upsampling is performed through pixel shuffle layers, followed by the .0of the high-resolution image I_{SR} through a final convolutional layer. This approach ensures that the low-resolution input is transformed into a high-quality super-resolved image.

3.2 Training Strategies

3.2.1 Knowledge Transfer Across Scales

To handle multiple scaling factors effectively, such as $\times 2$, $\times 3$, and $\times 4$, a progressive transfer learning strategy is used. Initially, the model is trained using a low scaling factor, specifically $\times 2$, which allows the network to learn the fundamental mappings required for picture super-resolution. Following the training of the model for the lower scaling factor, the learned weights are then used to initialize models for higher scaling factors, such as $\times 3$ and $\times 4$. This approach helps accelerate convergence when training for larger scaling factors the model can use knowledge acquired during previous stages of training.

3.2.2 Loss Function Analysis

The training process involves minimizing the pixel-wise reconstruction loss. Two primary loss functions are considered: L2 loss and L1 loss. The **L2 Loss** (Mean Squared Error), as defined in (4),endeavors to reduce the squared discrepancies between the forecasted and actual images, maximizing the PSNR. However, Although L2 loss helps to produce smoother results, it often tends to smooth out important details in the image.

$$\mathcal{L}L2 = \frac{1}{N} \sum i = 1^N \left(I_{SR}(i) - I_{HR}(i)\right)^2, \tag{4}$$

In contrast, the **L1 Loss** (Mean Absolute Error), given in (5), determines the overall discrepancy between the estimated and real pixel values. This loss function is preferred for image super-resolution tasks, as it preserves edges and textures, producing sharper and more detailed reconstructions compared to L2 loss.

$$\mathcal{L}L1 = \frac{1}{N} \sum i = 1^N \left|I_{SR}(i) - I_{HR}(i)\right|, \tag{5}$$

3.3 Multi-Scale Super-Resolution

Preprocessing and Evaluation Low-resolution (LR) images are generated by bicubic downsampling of high-resolution (HR) images at scaling factors $\times 2$, $\times 3$, and $\times 4$. During training, random patches of size 48×48 are extracted from both LR and HR images, with data augmentation applied through random rotations (90°, 180°, 270°) and horizontal or vertical flipping. The performance of the network is assessed on the DIV2K validation and test sets using PSNR and SSIM, which assess image quality and structural resemblance between the recovered and actual images.

Unified Framework and Training Process To handle different scaling factors, a unified framework with shared parameters across scales is employed. Each scaling factor has its own preprocessing module with residual blocks for normalization, while a shared main network extracts common features. Scale-specific modules perform the upsampling to generate high-resolution (HR) images. The bicubic downsampling for the low-resolution (LR) to HR mapping is

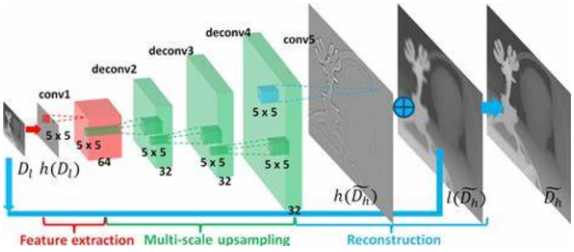

Figure 3: Multi-scale preprocessing working model(Hui et al., 2016)

mathematically expressed in (6). The network learns an inverse mapping to reconstruct I_{HR} from I_{LR}.

$$I_{LR} = \text{Downsample}(I_{HR}), \tag{6}$$

.

During training, a scaling factor is randomly selected ($s \in \{2,3,4\}$) for each mini-batch. The LR input is processed using scale-specific modules, passed through the shared network, and upsampled with scale-specific layers. L1 loss is computed, and weights are updated accordingly.

3.4 Implementation Approach

The model is built using PyTorch for its modular architecture and efficient GPU acceleration. It employs the L1 loss function for regression tasks and the Adam optimizer, configured with a learning rate of 10^{-4} for effective convergence. Data augmentation techniques, including random rotations and flips, increase dataset diversity, enhancing the model's adaptability to unfamiliar data and improving performance in real-world applications.

4 RESULTS AND DISCUSSION

In this section, we present and analyze the results of our proposed residual network-based framework for SISR. Performance is evaluated in terms of standard metrics such as PSNR and SSIM across multiple scales. Additionally, we compare it against leading approaches, including FSRCNN(Dong et al., 2016), Bicubic interpolation(Yuan et al., 2018), VDSR, Enhanced Deep SR(EDSR)(Lim et al., 2017), SR Generative Adversarial Network (SRGAN)(Ledig et al., 2017), and Cycle in Cycle GAN (CinCGAN)(Yuan et al., 2018).

4.1 Dataset: DIV2K

The DIV2K dataset, a standard for evaluating image super-resolution tasks, consists of higher quality 2K resolution images. For this study, the dataset is processed with a bicubic $\times 4$ degradation, generating low-resolution images at one-sixteenth the original size. The collection consists of 800 images designated for training, along with 100 images allocated for validation and an additional 100 images meant for testing, with the LR images serving as inputs and HR images as targets.

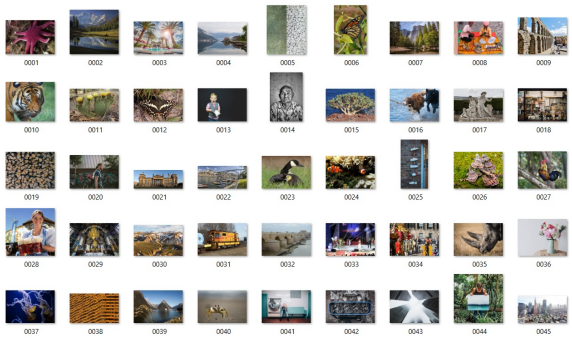

Figure 4: Sample images in the dataset.

4.2 Presentation of Results

The results for the $\times 4$ scaling evaluated on the DIV2K dataset are summarized in the table below. Two key metrics are used for evaluation: PSNR and SSIM.

The table presents a quantitative assessment of our scaling algorithm's performance on various high-resolution images from the DIV2K dataset. These results will help us understand the effectiveness of our approach and identify areas for improvement.

Table 1: Performance Comparison on DIV2K Validation Set for $\times 4$ Scaling (PSNR (dB) / SSIM)

Method	$\times 4$ **(PSNR)**	$\times 4$ **(SSIM)**
FSRCNN (CNN-Based)(Dong et al., 2016)	22.79	0.61
Bicubic(Yuan et al., 2018)	22.85	0.65
VDSR (Deep CNN)	28.17	0.65
EDSR (Deep CNN)(Lim et al., 2017)	22.67	0.62
SRGAN (GAN-Based)(Ledig et al., 2017)	24.33	0.67
CinCGAN (GAN-Based)(Yuan et al., 2018)	24.33	0.69
Proposed Model	**30.25**	**0.77**

4.3 Detailed Analysis of Results

Performance Highlights The proposed model demonstrates significant performance improvements over existing methods. It achieves the highest PSNR of 30.25 dB and SSIM of 0.77 for $\times 4$ scaling, outperforming both EDSR and SRGAN. Specifically, when compared to EDSR, the model improves PSNR by 7.57 dB and SSIM by 0.15, showcasing the effectiveness of the architectural optimizations imple-

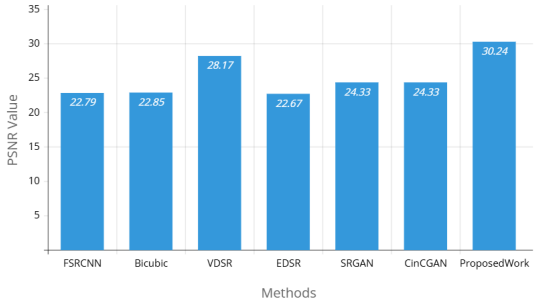

Figure 5: Result demonstrates proposed work is best concerning PSNR value

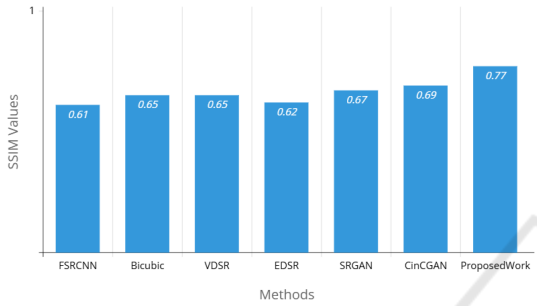

Figure 6: Result shows the proposed model is best concerning SSIM value

mented. While GAN-based methods such as SRGAN and CinCGAN provide competitive perceptual quality, they tend to introduce artifacts, leading to lower PSNR and SSIM values compared to the proposed model.

Comparison Observations In terms of comparison with other super-resolution techniques, FSRCNN and Bicubic interpolation, although lightweight and computationally efficient, struggle with recovering fine details due to their relatively simple architectures. VDSR, despite being a deeper model, achieves moderate improvements in PSNR and SSIM. SRGAN produces visually appealing results, but can occasionally introduce artifacts, negatively impacting the PSNR and SSIM scores. Similarly, CinCGAN performs well under specific degradation conditions, but lacks consistency when applied to general datasets, making it less reliable than the proposed method in a wider range of scenarios.

4.4 Qualitative Comparison

The proposed model shows notable improvements in image quality, offering sharper edges and better texture preservation compared to previous models such as FSRCNN and VDSR. It also exhibits fewer artifacts and superior perceptual quality relative to GAN-based approaches such as SRGAN. Furthermore, the

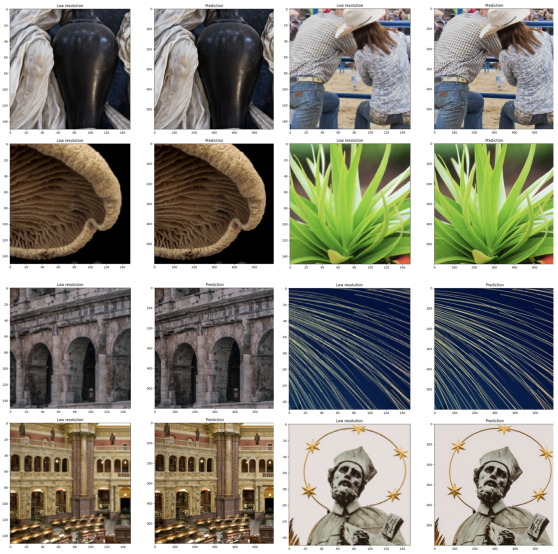

Figure 7: Super-resolution results of the proposed model for $\times 4$ scaling factors

model achieves significantly enhanced detail reconstruction over EDSR, indicating its effectiveness in fine-diameter refining while maintaining structural integrity.

4.5 Interpretation of Results and Implications

The experimental outcomes demonstrate the validity of the proposed residual network-based super-resolution model delivers superior reconstruction fidelity and structural quality for $\times 4$ scaling. These findings validate the original objective of achieving high-quality image reconstructions with an efficient and optimized network design. The practical implications of this work extend to various domains, including satellite images, medical imaging, and real-time video enhancement. Moreover, the model's scalability and adaptability to varying scaling factors ensure its usability across diverse applications without the need for separate models for each scale.

4.6 Limitations of the Current Approach

In spite of its cutting-edge capabilities, the suggested model has certain limitations. Firstly, the deep architecture necessitates significant computational resources, including GPU memory and processing power, which may pose challenges for deployment in resource-constrained environments. Secondly, the model's performance on real-world degra-

dation models, such as non-bicubic downsampling, may vary and require further adaptation. Finally, while the model achieves a balance between fidelity and perceptual quality, it may not fully replicate the aesthetic realism achieved by GAN-based approaches like SRGAN(Ledig et al., 2017) and CinCGAN(Yuan et al., 2018), which could be critical for applications prioritizing perceptual aesthetics.

5 CONCLUSION AND FUTURE WORK

Super-resolution using residual networks has greatly im proved high-resolution image reconstruction from low resolution inputs through residual learning and skip connec tions. This approach tackles issues like vanishing gradients and enables deeper architectures to capture fine details effectively. Achieving metrics such as a PSNR of 30.25 dB and SSIM of 0.77, the model outperforms traditional methods on the DIV2K dataset, making it suitable for precision-demanding applications like video enhancement. However, it still faces challenges, including high computational demands and adapt ing to real-world degradation, which need to be addressed for broader use in fields like medical imaging and consumer devices. Enhancing the balance between realism and struc tural fidelity could also improve its performance in aesthetic sensitive applications.

Integrating advanced techniques like self-attention, channel attention, and combining residual networks with vision trans formers can enhance performance. Innovations in lightweight designs, model pruning, and efficient training will aid deploy ment in resource-limited environments. Expanding datasets for diverse applications will strengthen the role of next-generation super-resolution technology.

REFERENCES

Aloysius, N. and Geetha, M. (2017). A review on deep convolutional neural networks. In *2017 international conference on communication and signal processing (ICCSP)*, pages 0588–0592. IEEE.

Bituin, R. C. and Antonio, R. (2024). Ensemble model of lanczos and bicubic interpolation with neural network and resampling for image enhancement. In *Proceedings of the 2024 7th International Conference on Software Engineering and Information Management*, pages 110–115.

Dong, C., Loy, C. C., and Tang, X. (2016). Accelerating the super-resolution convolutional neural network. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part II 14*, pages 391–407. Springer.

Hitawala, S., Li, Y., Wang, X., and Yang, D. (2018). Image super-resolution using vdsr-resnext and srcgan. *arXiv preprint arXiv:1810.05731*.

Hui, T.-W., Loy, C. C., and Tang, X. (2016). Depth map super-resolution by deep multi-scale guidance. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part III 14*, pages 353–369. Springer.

Khaledyan, D., Amirany, A., Jafari, K., Moaiyeri, M. H., Khuzani, A. Z., and Mashhadi, N. (2020). Low-cost implementation of bilinear and bicubic image interpolation for real-time image super-resolution. In *2020 IEEE Global Humanitarian Technology Conference (GHTC)*, pages 1–5. IEEE.

Kim, J., Lee, J. K., and Lee, K. M. (2016). Accurate image super-resolution using very deep convolutional networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1646–1654.

Kumar, S. (2020). V.: Perceptual image super-resolution using deep learning and super-resolution convolution neural networks (srcnn). *Intell. Syst. Comput. Technol*, 37(3).

Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z., et al. (2017). Photo-realistic single image super-resolution using a generative adversarial network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4681–4690.

Lim, B., Son, S., Kim, H., Nah, S., and Mu Lee, K. (2017). Enhanced deep residual networks for single image super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 136–144.

Luo, Z., Yu, J., and Liu, Z. (2019). The super-resolution reconstruction of sar image based on the improved fsrcnn. *The Journal of Engineering*, 2019(19):5975–5978.

Solutions, P. B. (2024). Low resolution vs high resolution images: Understanding the difference. Accessed: 30-December-2024.

Tian, C., Xu, Y., Zuo, W., Lin, C.-W., and Zhang, D. (2021). Asymmetric cnn for image superresolution. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 52(6):3718–3730.

Ullah, S. and Song, S.-H. (2023). Srresnet performance enhancement using patch inputs and partial convolution-based padding. *Computers, Materials & Continua*, 74(2).

Wang, Y., Rahman, S. S., and Arns, C. H. (2018). Super resolution reconstruction of μ-ct image of rock sample using neighbour embedding algorithm. *Physica A: Statistical Mechanics and its Applications*, 493:177–188.

Yang, C.-Y., Ma, C., and Yang, M.-H. (2014). Single-image super-resolution: A benchmark. In *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part IV 13*, pages 372–386. Springer.

Yang, J., Wright, J., Huang, T. S., and Ma, Y. (2010). Image super-resolution via sparse representation. *IEEE transactions on image processing*, 19(11):2861–2873.

Ye, S., Zhao, S., Hu, Y., and Xie, C. (2023). Single-image super-resolution challenges: a brief review. *Electronics*, 12(13):2975.

Yuan, Y., Liu, S., Zhang, J., Zhang, Y., Dong, C., and Lin, L. (2018). Unsupervised image super-resolution using cycle-in-cycle generative adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 701–710.

Yıldırım, M. and Dandıl, E. (2021). Automatic detection of multiple sclerosis lesions using mask r-cnn on magnetic resonance scans. *IET Image Processing*, 14.

Zhang, K., Sun, M., Han, T. X., Yuan, X., Guo, L., and Liu, T. (2017). Residual networks of residual networks: Multilevel residual networks. *IEEE Transactions on Circuits and Systems for Video Technology*, 28(6):1303–1314.