NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

MSAI

AI6121 Computer Vision

Paper Reading Report on

“Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network”

Ledig, C., et al (2017).

By:
Tiffany Tan Ge Ru
TIFF0030@e.ntu.edu.sg
G2404908D

Table of Contents

1 Introduction 3

1.1 Background 3

1.2 Motivation 3

1.3 Existing Work and Gaps 4

2 GANs and SRGAN 5

2.1 GANs 5

2.2 SRGANS 5

2.3 Methodology of SRGAN 5

3 Data, Training and Technique Evaluation 8

3.1 Data used 8

3.2 Training Process 8

3.3 Evaluation 9

4 Constraints 11

5 Related / Future Work 12

5.1 ESRGAN 12

5.2 Applications in the Logistics Industry 12

6 Conclusion 13

7 References 14

1 Introduction

1.1 Background

Working in the logistics industry - where surveillance systems like CCTVs in the warehouse tend to be using legacy systems with older models or environment is dimly lit, collecting quality image data for object detection and optical character recognition for computer vision projects has been one of my team's greatest challenges. Therefore, the paper I selected for this literature review was on the topic: **Super Resolution Generative Adversarial Network (SRGAN)**. It is a potential solution to increase the resolution of images without significant infrastructure changes and hence the cost involved. Written by Ledig et al. and titled **"Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network"**, the scope of the paper focuses on **Single Image Super Resolution (SISR)**. This report will discuss the motivations behind SRGAN, the paper's methodology, evaluation metrics, constraints and further discussions.

1.2 Motivation

Super-resolution (SR) involves generating high-resolution (HR) images from low-resolution (LR) counterparts. Before SRGAN, many super-resolution methods relied on minimising pixel-level errors, such as those measured by Mean Squared Error (MSE). However, these methods often produced blurry images and lacked the detailed textures necessary for realism.

The motivation for developing SR techniques like SRGAN arises from the need to improve image quality across various domains while overcoming hardware and resource limitations. Apart from the logistics industry, enhancing image quality is also crucial in fields such as medical imaging, forensic analysis and satellite observation, where high-resolution images are essential for detailed analysis and accurate diagnostics. SR techniques, including SRGAN, offer a cost-effective way to enhance lower-resolution images without the need for expensive high-resolution equipment.

SRGAN addresses the challenges posed by hardware limitations, where high-resolution cameras are expensive and difficult to maintain. By improving the resolution of images captured by lower-end devices, SRGAN helps conserve bandwidth in applications like video streaming, where lower-resolution images can be transmitted and then upscaled without sacrificing quality. For example, SRGAN can significantly enhance surveillance and security systems by improving the clarity of CCTV footage, making it easier to identify details in low-resolution videos.

Advances in AI also drive the development of SR techniques, as they contribute to the growth of generative models and inspire further innovations in image processing. The success of SRGAN has led to the development of models like ESRGAN, which improves upon SRGAN's foundation to produce even sharper images. I will be discussing ESRGAN in the future work section of this report.

1.3 Existing Work and Gaps

Traditional methods to tackle SISR include prediction-based ones such as linear, bicubic or Lanczos filtering. While computationally efficient and straightforward, the tradeoff was that they tend to yield results with overly smooth textures. The traditional approaches tend to focus on minimizing pixel-wise errors, typically using Mean Squared Error (MSE) as the loss function.

Other SISR methods include Example-based and self-similarity methods, like those proposed by Glasner et al. (2009) [8], which exploit the redundancy of patches within and across scales in the image. By finding similar patches within the image itself or using external databases, these methods reconstruct HR images by blending the corresponding HR patches. However, the effectiveness of these methods is limited by their reliance on finding good patch matches, which can be sparse or noisy, leading to artifacts or blurred regions [8].

Regression-based methods, including Random Forests (Schulter et al., 2015) [11] and Gaussian Process Regression (He and Siu, 2011) [12], offer another approach by predicting the HR image from the LR image using learned regression models. These methods have shown good results but are often constrained by the capacity of the regression models, requiring large amounts of training data and being prone to overfitting, particularly when the training data is not representative of the testing scenarios (Schulter et al., 2015) [11].

Early deep learning approaches, like SRCNN (Dong et al., 2016) [13], marked a departure by using deep networks to learn the relationship between low-resolution (LR) and high-resolution (HR) images. While these models, including VDSR and DRCN (Kim et al., 2016) [18], offered improvements, they still resulted in images that lacked the fine textures necessary for perceptual quality.

Efforts to address these limitations have led to the development of edge-preservation methods and more advanced techniques, such as convolutional sparse coding (Gu et al., 2015) [19] and further exploitation of self-similarity [8]. However, the primary gap in existing SR methods remains their focus on pixel-level accuracy at the expense of perceptual realism. Even the most advanced learning-based methods often fall short of capturing the nuanced textures and details essential for producing visually convincing HR images.

Therefore, gave rise to the birth of SRGAN, which was to bridge this gap by generating images that not only meet quantitative metrics like Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM) but also align with human visual perception. SRGAN introduced the use of perceptual loss, focusing on feature-space similarity rather than pixel-space, and combined this with adversarial loss from the GAN framework to produce visually convincing, as evidenced by higher Mean Opinion Scores (MOS) in human evaluations. This shift from pixel accuracy to perceptual quality represents a significant evolution in SISR, addressing the critical gaps left by earlier methods.

2 GANs and SRGAN

2.1 GANs

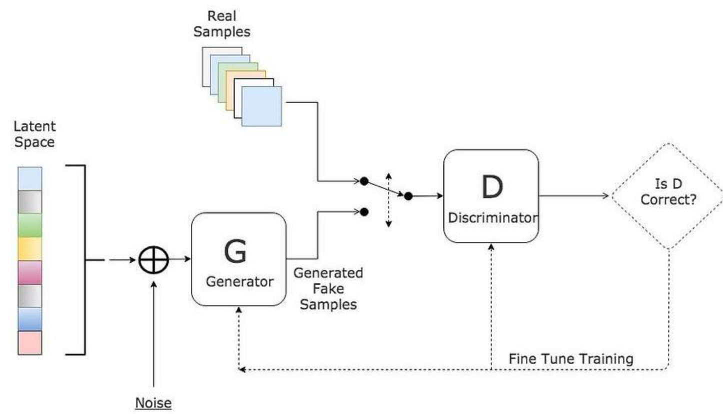

Figure 1: General GAN architecture [15]

As shown in Figure 1 above, GANs (Generative Adversarial Network) consists of two networks – a Generator and Discriminator Network. Initially introduced by Goodfellow [2], the generator attempts to generate real images that are indistinguishable from fake images, while the discriminator tries to differentiate the two. It is essentially a min-max problem, where the discriminator tries to minimise the loss function while generator tries to do the opposite, which is to maximise the loss function. Training is complete when the generator has successfully recreated images that are indistinguishable from real ones.

2.2 SRGANS

SRGANs (Super-Resolution Generative Adversarial Network) is a specific type of GAN, designed to for the task of single image resolution. SRGAN applies a deep network in conjunction with an adversarial network to generate high-resolution images from low-resolution inputs. It can generate high-resolution images that look more realistic and visually appealing compared to traditional methods. The architecture consists of two primary components: 1) a generator network, which is a deep convolutional network designed to create high-resolution images from low-resolution inputs, and 2) a discriminator network, which differentiates between real high-resolution images and those outputs from the generator.

2.3 Methodology of SRGAN

The goal is to map a low-resolution image I^{LR} to a high-resolution I^{HR}, achieved through training a generator network as a feedforward CNN G_{θ_G}. The optimization is driven by a perceptual loss function, which is a composition of: content loss and adversarial loss - tailored to balance structural fidelity and perceptual realism. The formula is as below:

$$l^{SR} = \underbrace{\underbrace{l^{SR}_{\mathrm{X}}}_{\text{content loss}} + \underbrace{10^{-3} l^{SR}_{Gen}}_{\text{adversarial loss}}}_{\text{perceptual loss (for VGG based content losses)}}$$

The content loss is computed using feature representations from a pre-trained VGG-19 network, as described by Simonyan & Zisserman (2015) [10]. This network has been widely adopted in tasks requiring feature-based similarity assessments due to its capacity to capture rich hierarchical representations of image content, as demonstrated by Johnson et al. (2016) [17] and Bruna et al. (2015) [16]. Instead of minimizing pixel-wise discrepancies, which often results in overly smooth outputs lacking detail [7], the content loss compares feature activations at various layers of the VGG-19 network, thereby preserving crucial structural information such as edges, textures, and spatial hierarchies in the generated images.

In parallel, the adversarial loss component is defined within a Generative Adversarial Network (GAN) framework. Here, a discriminator network D_{θ_D} is trained alongside the generator to solve an adversarial min-max problem.

$$\min_{\theta_G} \max_{\theta_D} \mathbb{E}_{I^{HR} \sim p_{\text{train}}(I^{HR})}[\log D_{\theta_D}(I^{HR})] + \mathbb{E}_{I^{LR} \sim p_G(I^{LR})}[\log(1 - D_{\theta_D}(G_{\theta_G}(I^{LR}))]$$

The discriminator is tasked with distinguishing real high-resolution images from those generated by G_{θ_G}, while the generator aims to produce images that can deceive the discriminator. The architecture is as shown in Figure 1 below.

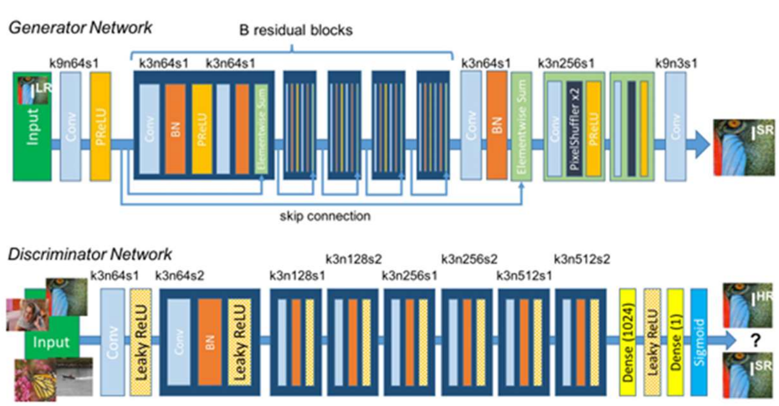

Figure 2: Architecture of Generator and Discriminator Network with corresponding kernel size (k), number of feature maps (n) and stride (s) indicated for each convolutional layer [1]

2.2.1 Generator Network G

The architecture of the generator comprises multiple residual blocks, inspired by the work of He et al. (2016) [3], where each block includes two convolutional layers with 3x3 kernels, followed by batch normalization and Parametric ReLU (PReLU) activation. This sequence is repeated within each residual block.

After processing in a residual block, an elementwise sum is performed with the input to the block, and the output is passed to the next block, repeating the process across all 16 residual blocks, as specified in the original SRGAN design. These blocks facilitate the training of deeper networks by allowing gradients to propagate more effectively, thus enabling the learning of more complex features.

Once the residual blocks are processed, the output undergoes another convolutional and batch normalization layer. The output of the first parametric ReLU function is again subjected to an elementwise sum. As proposed by Shi et al. (2016) [4], the up-sampling block uses pixel shuffling to gradually increase image resolution. This upscaling consists of two blocks, and the process concludes with a final convolutional layer, producing the super-resolution image as the output.

2.2.1 Discriminator Network D

A discriminator network is essentially an image classification Convolution Neural Network (CNN). Its role is to differentiate generated and high-resolution images by first learning to classify the image. The discriminator network, structured in accordance with the guidelines set forth by Radford et al. (2016) [6] in their DCGAN architecture, consists of eight convolutional layers with progressively increasing numbers of filters. As the network progresses, the spatial size of the feature map is intentionally reduced using strided convolutions, while increasing the depth (number of feature maps). This allows the network to capture more complex patterns and relationships within the image at multiple scales and helps the discriminator more efficiently extract and process features. LeakyReLU is used as the activation function ($\alpha = 2$) to maintain a small enough gradient for negative inputs to mitigate issues of vanishing gradients during training.

The resulting high-level feature maps are passed through dense layers and a sigmoid function to produce a final probability, enabling the discriminator to classify images as real or generated.

This adversarial setup enables the generator to learn to produce outputs that can fool the Discriminator D that was trained to distinguish real from super-resolved images. Generator can hence generate images where 1) the structural details captured by the content loss and 2) align with the manifold of natural images, resulting in perceptually realistic high-resolution images.

3 Data, Training and Technique Evaluation

3.1 Data used

In this paper, experiments were performed on data from three public datasets, comprising of 350,000 images with a scale factor 4x between low- and high-resolution images. This corresponds to starting with images that had 16x times fewer pixels than the high resolutions photos to be created.

To ensure a fair comparison, images were cropped, and a small border was removed to avoid any edge effects. The paper also measured the quality of images using two metrics, PSNR (Peak Signal to Noise Ratio) and SSIM (Structural Similarity Index), focusing on the y-channel (brightness) of the images.

Results from other methods like nearest neighbour, bicubic, SRCNN and SelfExSR were used for cross validation of results. A statistical test called the Wilcoxon signed-rank test to confirm the significance of results, considering a significance level of $p < 0.05$.

3.2 Training Process

For pre-processing, the low-resolution (LR) images were scaled to a range of [0,1], while the high-resolution (HR) images were scaled to [-1,1] by first normalizing the pixel values from [0,255] to [0,1] and then adjusting them to the desired range. This scaling ensures that different features contribute equally to the learning process, facilitating better gradient flow and faster convergence during training.

Before training the full SRGAN, the generator network was pre-trained as an SRResNet using Mean Squared Error (MSE) loss, a standard loss function for image reconstruction tasks that minimizes the average squared difference between the predicted and ground truth high-resolution images. To maintain consistency in the loss calculation, VGG feature maps used in the perceptual loss were rescaled to match the scale of the MSE loss.

For optimization, the Adam algorithm was used with β=0.9 and applied with a learning rate of 0.0001 over 1 million update iterations for the SRResNet. This pre-trained SRResNet was then used to initialize the generator for SRGAN training, which helps avoid poor local minima during the adversarial training phase. The SRGAN variants were further trained for 100,000 iterations at a learning rate of 0.0001, followed by another 100,000 iterations at a reduced learning rate of 0.00001. During training, updates were alternated between the generator and discriminator networks, a standard technique in GAN training. The generator network consisted of 16 identical residual blocks.

To ensure consistent output during testing, batch normalization updates were disabled, following practices consistent with the Theano and Lasagne frameworks.

3.3 Evaluation

SRGAN was evaluated based on Mean Opinion Score Testing (MOS) and the effect of different content loss.

3.3.1 Mean Opinion Score Testing

This paper utilized MOS testing to assess how well different super-resolution methods produced perceptually convincing images. This is a commonly used subjective quality measuring tool where each observer has the expertise to rank the photos from 1 to 5 (best, i.e. very good high-resolution images) [9]. The results showed that SRGAN variants, particularly those using higher-level VGG features for content loss, produced more visually convincing images compared to other methods. The MOS scores confirmed that SRGAN generally outperformed other techniques in terms of perceptual quality.

3.3.2 Content Loss evaluation

Next, the effect of different content loss choices in the perceptual loss for GAN-based networks was evaluated. Losses based on standard MSE and on feature maps extracted from different layers of a VGG network were compared. Firstly, for SRGAN-MSE, which is to investigate the adversarial network with standard MSE as content loss, results produced the highest PSNR values, indicating good pixel-level accuracy but lacking perceptual accuracy.

Moreover, the study found that SRGAN using higher-level VGG features (VGG54) achieved better texture details compared to those using lower-level features (VGG22). This finding underscores the importance of selecting appropriate content loss functions to enhance perceptual quality. The trend observed was that higher-level VGG feature maps lead to improved texture details. As a result, SRGAN-VGG54 was chosen as the SRGAN variant to be further compared with other state-of-the-art methods.

The writer hypothesizes that deeper layers of the VGG network, when used for content loss, concentrate on preserving the core content of an image, while the adversarial loss enhances texture details, leading to perceptually the most convincing results. This dual approach helps overcome the over smoothing seen in images that rely solely on content loss.

3.3.3 Overall Performance

Table 1: Comparison of NN, bicubic, SRCNN, SelfExSR, DRCN, ESPCN, SRResNet, SRGAN-VGG54 [1]

Set5	nearest	bicubic	SRCNN	SelfExSR	DRCN	ESPCN	**SRResNet**	**SRGAN**	HR
PSNR	26.26	28.43	30.07	30.33	31.52	30.76	**32.05**	29.40	∞
SSIM	0.7552	0.8211	0.8627	0.872	0.8938	0.8784	**0.9019**	0.8472	1
MOS	1.28	1.97	2.57	2.65	3.26	2.89	3.37	**3.58**	4.32
Set14									
PSNR	24.64	25.99	27.18	27.45	28.02	27.66	**28.49**	26.02	∞
SSIM	0.7100	0.7486	0.7861	0.7972	0.8074	0.8004	**0.8184**	0.7397	1
MOS	1.20	1.80	2.26	2.34	2.84	2.52	2.98	**3.72**	4.32
BSD100									
PSNR	25.02	25.94	26.68	26.83	27.21	27.02	**27.58**	25.16	∞
SSIM	0.6606	0.6935	0.7291	0.7387	0.7493	0.7442	**0.7620**	0.6688	1
MOS	1.11	1.47	1.87	1.89	2.12	2.01	2.29	**3.56**	4.46

The results, summarised in Table 1 above, demonstrate that SRResNet sets a good benchmark in terms of PSNR and SSIM across all three datasets. These results highlight its strong performance in preserving pixel-level accuracy. However, when assessing perceptual quality through MOS, SRGAN consistently outperformed other methods. SRGAN achieved MOS scores of 3.58 on Set5, 3.72 on Set14, and 3.56 on BSD100, surpassing all other methods including SRResNet. This suggests that SRResNet excels in objective metrics like PSNR and SSIM, while SRGAN is superior in generating images that are more visually appealing to the human eye. Qualitative results can be seen in Figure 3 below.

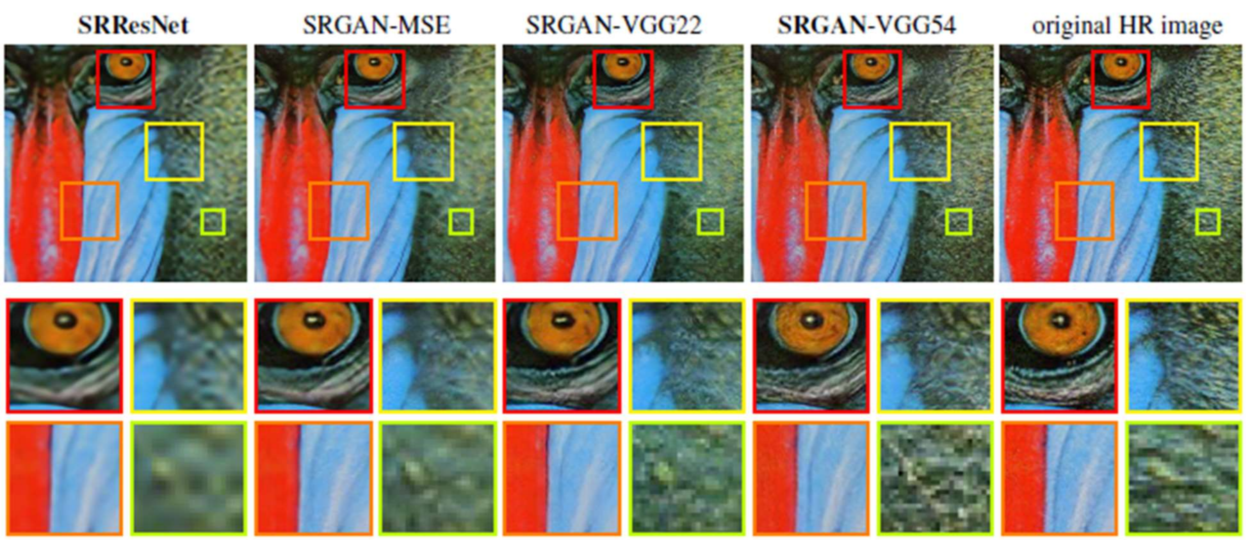

Figure 3: Reconstruction results from SRResNet, SRGAN-MSE, SRGAN-VGG22, SRGAN-VGG54 and corresponding reference high-resolution image [1]

Thus, this paper has confirmed that SRGAN reconstructions for large upscaling factors (4x) are more photorealistic than existing methods.

4 Constraints

Despite its advancements, SRGAN does have its limitations.

First, SRGAN requires significant computational resources both during training and inference. The deep architectures used in SRGAN, which are necessary for capturing complex image features, result in longer training times and higher GPU usage, making it less feasible for real-time or resource-constrained applications (Ledig et al., 2017) [1].

Another challenge is the inherent instability in training GANs, where the balance between the generator and discriminator must be carefully managed to prevent issues such as mode collapse or non-convergence [2]. SRGAN is particularly sensitive to hyperparameter choices, which can lead to considerable variation in output quality, making it difficult to generalize across different datasets or applications [5].

Artifact generation is also a notable issue with SRGAN. While it excels in enhancing texture details, this can sometimes result in high-frequency artifacts, especially when the adversarial loss overly emphasizes texture generation. These artifacts can detract from the realism of the generated images, which is particularly problematic in applications requiring high visual accuracy, such as medical imaging or scientific analysis (Blau & Michaeli, 2018) [20].

5 Related / Future Work

5.1 ESRGAN

In 2018, Xintao et al. introduced the Enhanced Super-Resolution Generative Adversarial Network (ESRGAN), marking a significant advancement over SRGAN. ESRGAN replaces the traditional residual block with the Residual-in-Residual Dense Block (RRDB), which improves network stability and allows for deeper architectures without the need for batch normalization, thereby reducing artifacts. Additionally, ESRGAN employs a Relativistic Average GAN (RaGAN) instead of the standard GAN approach, enhancing the discriminator's ability to compare the realism of generated images relative to real ones. The refinement of perceptual loss, using feature maps from earlier layers of the VGG network, further enhances the visual quality of the images, particularly in terms of edge sharpness and brightness consistency. The evolution from SRGAN to ESRGAN underscores the dynamic nature of the field and the ongoing efforts to push the boundaries of super-resolution technology [1,14].

5.2 Applications in the Logistics Industry

Being a data scientist in the logistics industry, I explored this paper for the purpose of understanding how it could value add to my work. One significant area of impact is enhanced image quality for surveillance. In large-scale warehouses, obtaining high-quality video surveillance is critical for ensuring security and optimising operational workflows. SRGAN and ESRGAN can be employed to increase the resolution of video feeds from surveillance cameras, particularly those capturing low-resolution images due to bandwidth limitations or outdated camera technology. By improving the clarity and detail of these images, these models facilitate more precise monitoring and identification of objects, such as tracking the movement of goods, or identifying issues like damaged packaging. This improved visual data is crucial for maintaining security standards and operational efficiency in complex logistics environments.

Another area where SRGAN/ESRGAN can significantly contribute is in optimising automated systems within warehouses. Automated systems, including robotic pickers and Automated Storage and Retrieval Systems (AS/RS), rely on high-resolution images to accurately identify and handle items. By enhancing the resolution and detail of images used by these systems, SRGAN/ESRGAN help reduce errors in picking and sorting processes, thereby improving the overall efficiency of warehouse operations. Additionally, in environments where Optical Character Recognition (OCR) is used to read labels, barcodes, or QR codes on packages, the improved image resolution provided by these models enhances the accuracy and speed of OCR readings. This, in turn, leads to more efficient sorting and tracking processes, which will reduce manhours wasted due to human error or searching for missing shipments.

By addressing the limitations of low-resolution imaging in these critical areas, SRGAN/ESRGAN can play a pivotal role in advancing the technological capabilities of the logistics industry, supporting more secure, accurate, and efficient warehouse management systems.

6 Conclusion

SRGAN has significantly advanced super-resolution in computer vision by combining pixel accuracy with perceptual quality, using GANs to generate high-resolution images that closely match human visual perception. This is important as the goal of computer vision is to mimic human vision and how we perceive and interpret visual information. This innovation has broad applications, from improving medical imaging to enhancing surveillance and optimising automated systems in logistics. Building on SRGAN's foundation, ESRGAN addresses its limitations, further refining image realism. These advancements highlight the continuing impact of SRGAN and its successors on the future of computer vision, driving forward improvements in image processing across various fields. As a data scientist, I am excited to explore these techniques to improve the accuracy and precision of my work.

7 References

[1] Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z., & Shi, W. (2017). *Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network*. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Retrieved from https://arxiv.org/pdf/1609.04802

[2] Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z., & Shi, W. (2017). Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network. arXiv preprint arXiv:1609.04802. https://arxiv.org/abs/1609.04802

[3] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 770-778. https://doi.org/10.1109/CVPR.2016.90

[4] Shi, W., Caballero, J., Huszár, F., Totz, J., Aitken, A. P., Bishop, R., Rueckert, D., & Wang, Z. (2016). Real-Time Single Image and Video Super-Resolution Using an Efficient Sub-Pixel Convolutional Neural Network. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 1874-1883. https://doi.org/10.1109/CVPR.2016.207

[5] Sajjadi, M. S. M., Schölkopf, B., & Hirsch, M. (2017). EnhanceNet: Single image super-resolution through automated texture synthesis. Proceedings of the IEEE International Conference on Computer Vision (ICCV), 4491-4500. https://doi.org/10.1109/ICCV.2017.480

[6] Radford, A., Metz, L., & Chintala, S. (2016). Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv preprint arXiv:1511.06434. https://arxiv.org/abs/1511.06434

[7] Mathieu, M., Couprie, C., & LeCun, Y. (2015). Deep multi-scale video prediction beyond mean square error. arXiv preprint arXiv:1511.05440. https://arxiv.org/abs/1511.05440

[8] Glasner, D., Bagon, S., & Irani, M. (2009). Super-resolution from a single image. 2009 IEEE 12th International Conference on Computer Vision, 349-356. https://doi.org/10.1109/ICCV.2009.5459271

[9] Lepcha, D. C., Goyal, B., Dogra, A., & Goyal, V. (2023). Image super-resolution: A comprehensive review, recent trends, challenges and applications. Information Fusion, 91, 230-260. https://doi.org/10.1016/j.inffus.2022.10.004

[10] Simonyan, K., & Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. In Proceedings of the International Conference on Learning Representations (ICLR). Retrieved from https://arxiv.org/abs/1409.1556

[11] Schulter, S., Leistner, C., & Bischof, H. (2015). Fast and accurate image upscaling with super-resolution forests. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 3791-3799. https://doi.org/10.1109/CVPR.2015.7299003

[12] He, H., & Siu, W. C. (2011). Single image super-resolution using Gaussian process regression. *2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 449-456. https://ieeexplore.ieee.org/document/5995527

[13] Dong, C., Loy, C. C., He, K., & Tang, X. (2016). Image super-resolution using deep convolutional networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 38(2), 295-307. https://ieeexplore.ieee.org/document/7115171

[14] Wang, X., Yu, K., Wu, S., Gu, J., Liu, Y., Dong, C., Loy, C. C., Qiao, Y., & Tang, X. (2018). ESRGAN: Enhanced super-resolution generative adversarial networks. arXiv preprint arXiv:1809.00219. https://arxiv.org/abs/1809.00219

[15] Analytics Vidhya. (2020, March 17). *General GAN architecture* [Diagram]. In A review of Generative Adversarial Networks (Part 1). Medium. https://medium.com/analytics-vidhya/a-review-of-generative-adversarial-networks-part-1-a3e5757a3dc2

[16] Johnson, J., Alahi, A., & Fei-Fei, L. (2016). Perceptual Losses for Real-Time Style Transfer and Super-Resolution. In Proceedings of the European Conference on Computer Vision (ECCV). arXiv:1603.08155. Retrieved from https://arxiv.org/pdf/1603.08155

[17] Bruna, J., Sprechmann, P., & LeCun, Y. (2015). Super-Resolution with Deep Convolutional Sufficient Statistics. arXiv:1511.05666. Retrieved from https://arxiv.org/pdf/1511.05666

[18] Kim, J., Lee, J. K., & Lee, K. M. (2016). Deeply-recursive convolutional network for image super-resolution. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 1637-1645. https://arxiv.org/abs/1511.04491

[19] Gu, S., Zhang, L., Zuo, W., & Feng, X. (2015). Convolutional sparse coding for image super-resolution. 2015 IEEE International Conference on Computer Vision (ICCV), 1823-1831. https://doi.org/10.1109/ICCV.2015.212

[20] Blau, Y., & Michaeli, T. (2018). The Perception-Distortion Tradeoff. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 6228-6237. https://doi.org/10.1109/CVPR.2018.00652

[21]* OpenAI. *ChatGPT*. OpenAI, September 24, 2024, chat.openai.com.

*ChatGPT was used to review for grammatical correctness and semantic structure of texts.