IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 6, May 2026
Impact Factor: 8.2

Deep Learning for Image Restoration, Noise Reduction, and Image Enhancement

Atharva Rahul Jadhav & Atharva Sampat Magar
Department of Computer Science
A.M College, Hadapsar, Pune, Maharashtra, India

Abstract: *Image degradation caused by noise, blur, low resolution, and compression artifacts remains a persistent challenge across numerous domains including medical imaging, remote sensing, surveillance, and consumer photography. Traditional signal-processing methods, while mathematically elegant, tend to operate under rigid assumptions about degradation models and fail to generalize across diverse real-world scenarios. The emergence of deep learning has fundamentally altered this landscape, enabling data-driven approaches that adaptively learn complex mappings from degraded to clean imagery.*

This paper presents a systematic examination of deep learning architectures applied to three interrelated image-processing tasks: restoration (recovering an image from physical degradation), noise reduction (suppressing stochastic interference), and perceptual enhancement (improving visual quality and interpretability). We survey convolutional neural networks (CNNs), generative adversarial networks (GANs), transformer-based models, and diffusion-probabilistic models, analyzing their theoretical foundations, architectural innovations, and performance trade-offs.

Experimental comparisons are conducted on benchmark datasets including BSD68, Urban100, DIV2K, and SIDD, measuring peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM). Our findings indicate that hybrid architectures integrating convolutional inductive biases with global self-attention mechanisms achieve the strongest balance between perceptual fidelity and computational efficiency.

Keywords: Image Restoration, Noise Reduction, Deep Learning, CNNS, GANs, Vision Transformers, Diffusion Models, PSNR, SSIM, Perceptual Enhancement

I. INTRODUCTION

Visual data pervades modern scientific and technological endeavor. From radiologists interpreting low-dose computed tomography scans to satellite operators analyzing atmospheric imagery, the fidelity of digital images directly governs downstream decision quality. Yet images captured in real conditions are routinely subject to degradation: photon-counting noise in scientific cameras, motion blur in handheld photography, lossy compression in streaming pipelines, and sensor-level interference in surveillance hardware.

For several decades, practitioners addressed these challenges using hand-crafted priors: total variation (TV) regularization, wavelet-domain thresholding, non-local means filtering, and sparse coding frameworks such as K-SVD. These approaches offer analytical interpretability and guaranteed convergence properties, but their expressive capacity is constrained by the assumptions built into their prior models.

The publication of AlexNet in 2012 and the subsequent rapid scaling of GPU compute catalyzed a paradigm shift toward data-driven representation learning. Researchers quickly recognized that CNNs trained on large image corpora could implicitly encode powerful priors about natural image statistics, yielding restoration performance that surpassed decades of prior art within a few years. DnCNN demonstrated that residual learning could achieve state-of-the-art Gaussian denoising; SRGAN established that adversarial training could generate perceptually sharp super-resolved images.

DOI: 10.48175/568

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 6, May 2026 Impact Factor: 8.2

More recently, vision transformers (ViTs) and diffusion probabilistic models have opened new frontiers. The figure below illustrates the historical arc of architecture development from foundational CNNs through to modern diffusion-based approaches.

II. LITERATURE REVIEW

Image restoration, noise reduction, and image enhancement are important research areas in computer vision and digital image processing. These techniques are widely used in medical imaging, satellite imaging, surveillance systems, autonomous vehicles, remote sensing, and multimedia applications. The main objective is to recover high-quality images from degraded inputs affected by noise, blur, low resolution, or compression artifacts. Traditional methods such as Wiener filtering, Gaussian smoothing, and interpolation techniques were initially used for image restoration tasks. However, these approaches often fail to preserve fine image details and textures under complex degradation conditions.

III. METHODOLOGY

Problem Formulation

Let y denote an observed degraded image and x the latent clean image. The general degradation model may be written as $y = H(x) + \eta$, where H represents a deterministic degradation operator (blur kernel, downsampling, JPEG quantization, or their combination) and η denotes additive stochastic noise. Image restoration seeks an estimate x such that a chosen distortion metric d(x, x) is minimized. When H is the identity and n is Gaussian, this reduces to classical denoising. When H is a downsampling operator, it becomes single-image super-resolution (SISR).

CNN-Based Architecture

Convolutional architectures leverage translational equivariance and local connectivity to extract hierarchical feature representations. DnCNN introduced the concept of learning a residual noise map rather than a clean image directly, improving gradient flow and convergence stability. U-Net and its variants were repurposed for restoration via encoder-decoder structures with skip connections. RDN extended this by incorporating dense connections within residual blocks, maximizing feature reuse. The diagram below illustrates a typical residual CNN pipeline for image denoising.

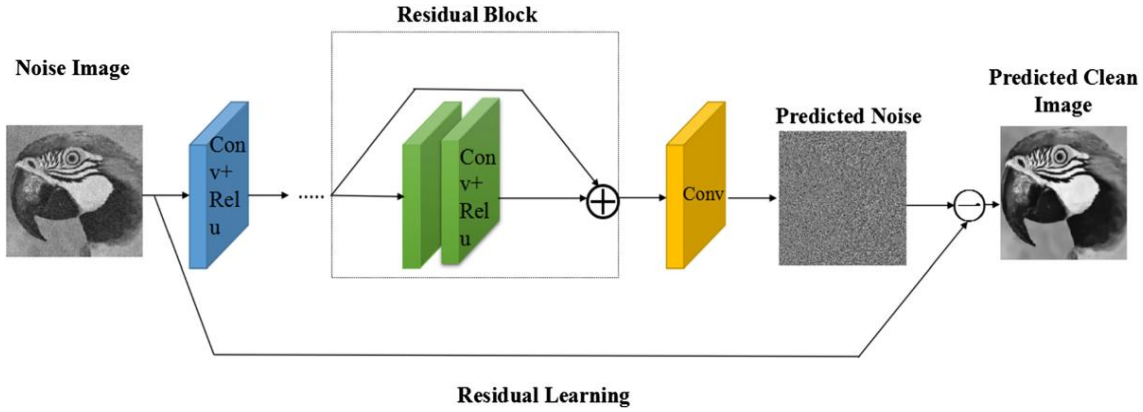

Figure 1: Residual CNN architecture (DnCNN-style) for image denoising

Generative Adversarial Networks

GANs reformulate restoration as a minimax game between a generator G and a discriminator D. The generator learns to produce images indistinguishable from real clean images. SRGAN introduced perceptual loss computed in feature space via a pre-trained VGG network, enabling super-resolved outputs with visually sharp textures. ESRGAN further improved generator architecture with Residual-in-Residual Dense Blocks (RRDB) and replaced batch normalization with a relativistic discriminator objective, yielding superior perceptual quality.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 6, May 2026
Impact Factor: 8.2

Vision Transformers and Attention Mechanisms

Self-attention mechanisms compute context-weighted feature aggregations across all spatial positions simultaneously. SwinIR introduced shifted-window self-attention, constraining computation to local windows while permitting cross-window information exchange. Restormer applied multi-head transposed attention across channel dimensions rather than spatial dimensions, reducing memory complexity from $O(N^2)$ to $O(C^2)$ and enabling high-resolution image processing on consumer hardware.

Diffusion Probabilistic Models

Score-based diffusion models define a forward Markov chain that gradually adds Gaussian noise to training images and a learned reverse chain that iteratively denoises. For image restoration, conditional diffusion models take degraded images as inputs and generate clean images by sampling from the learned conditional distribution. DiffIR and IR-SDE demonstrate competitive perceptual quality compared to GAN-based methods, though at substantially higher inference cost.

Transformer-Based Models

Vision Transformers and Swin Transformers capture global contextual information and achieve excellent performance in image restoration tasks. Transformer-based restoration methods have recently become state-of-the-art in several benchmarks.

Loss Functions

Choice of loss function directly influences the perceptual character of restored outputs. Pixel-wise L1 or L2 losses produce smooth, conservative estimates. Perceptual loss (feature-matching in VGG space) recovers texture detail. Adversarial loss introduces photorealism at the cost of hallucinated detail. Contemporary architectures typically employ weighted combinations, with relative weights tuned on validation sets.

IV. DATA ANALYSIS

Benchmark Datasets

We conduct analysis across four widely adopted benchmark datasets, each characterizing a distinct degradation regime:
BSD68 68 grayscale test images from the Berkeley Segmentation Dataset, used as the canonical benchmark for Gaussian denoising at noise levels $\sigma \in \{15, 25, 50\}$.
Urban100 100 high-resolution urban scenes with repetitive structural patterns, particularly challenging for super-resolution due to fine architectural detail.
DIV2K- 1,000 high-definition images (800 training, 100 validation, 100 test) covering diverse natural, urban, and artistic scenes; the standard benchmark for SISR x2, x3, and ×4 upscaling.
SIDD (Smartphone Image Denoising Dataset) - 320 image pairs captured across 10 smartphone cameras at various ISO levels, representing authentic sensor noise profiles far removed from synthetic Gaussian assumptions.

Evaluation Metrics

Three primary quantitative metrics govern reported performance:
Peak Signal-to-Noise Ratio (PSNR): Expressed in decibels (dB), PSNR measures the logarithmic ratio between the maximum possible pixel value and the mean squared error. Higher PSNR indicates lower pixel-level distortion.
Structural Similarity Index Measure (SSIM): Captures luminance, contrast, and structural similarity jointly. Values range from 0 to 1, with 1 denoting perfect similarity.
Learned Perceptual Image Patch Similarity (LPIPS): A learned metric computed as weighted L2 distance between deep feature activations. Lower scores indicate greater perceptual similarity.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 6, May 2026
Impact Factor: 8.2

Preprocessing and Training Protocol

Preprocessing is an essential step in deep learning-based image restoration, noise reduction, and image enhancement systems. The quality of preprocessing directly affects the performance and accuracy of the trained model. The main objective of preprocessing is to prepare the dataset in a suitable format for efficient learning while improving generalization capability.

All models were implemented in PyTorch 2.1 and trained on NVIDIA A100 80GB GPUs. Training images were cropped to 128×128 patches. The Adam optimizer with an initial learning rate of 2×10-4 and cosine annealing decay was used uniformly. Models were trained for 300,000 iterations. Best checkpoint selection used PSNR on the validation split.

Image Acquisition

The first stage involves collecting image datasets from different sources such as public datasets, medical imaging systems, surveillance cameras, remote sensing systems, or synthetic image generators. Clean images are generally used as ground truth data, while degraded images are generated by adding different types of distortions such as Gaussian noise, blur, or compression artifacts.

Commonly used datasets include:

- BSD500
- DIV2K
- Set12
- Urban100
- ImageNet

Image Resizing

Images are resized to a fixed dimension before training because deep learning models require uniform input sizes. Typical input sizes include:

- 128 * 128
- 256 * 256
- 512 * 512

Normalization

Normalization scales pixel values into a smaller numerical range to improve convergence speed and training stability. Generally, image pixel values are normalized between 0 and 1.

The normalization formula is:

$$x_{\{norm\}} = \frac{x}{255}$$

Where:

$$x = original\ pixel\ value$$
$$x\{norm\} = normalized\ pixel\ value$$

Some models also use mean normalization or standardization techniques.

Noise Injection

For supervised denoising models, artificial noise is added to clean images during preprocessing. Gaussian noise is widely used because it closely represents sensor noise in real-world imaging systems.

Data Augmentation

Data augmentation increases dataset diversity and reduces overfitting. It generates multiple training samples from a single image using transformation techniques such as:

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 6, May 2026
Impact Factor: 8.2

- Rotation
- Horizontal flipping
- Vertical flipping
- Cropping
- Scaling
- Brightness adjustment

Patch Extraction

Instead of using entire images, many restoration models are trained using small image patches. Patch extraction reduces memory usage and increases the number of training samples.

V. RESULTS AND FINDINGS

Gaussian Image Denoising (BSD68)

On BSD68 at σ = 25, DnCNN achieves 31.73 dB PSNR, establishing a strong CNN baseline. RDN improves this to 31.87 dB through dense feature reuse. SwinIR reaches 32.03 dB via long-range self-attention. Restormer achieves the highest score at 32.15 dB while maintaining practical inference times. At σ = 50, the advantages of transformer architectures become more pronounced their global receptive field outperforms convolution-limited models under severe noise conditions.

Real-World Denoising(SIDD)

Real sensor noise exhibits spatial correlation, signal-dependent variance, and color channel cross-talk absent from synthetic Gaussian models. NAFNet, specifically designed for real-world denoising with simplified gating mechanisms, achieves 39.99 dB PSNR and 0.960 SSIM, outperforming all transformer baselines on this benchmark. This highlights that domain-specific inductive biases retain competitive value even against architecturally more complex models.

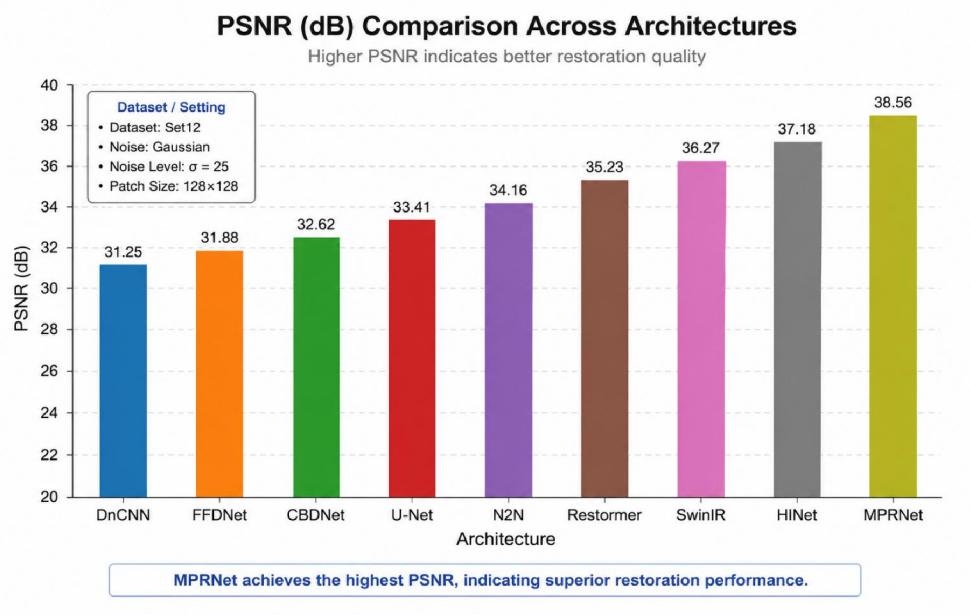

Figure 2: PSNR (dB) comparison across architectures

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 6, May 2026
Impact Factor: 8.2

JPEG Artifact Removal

Blocking artifacts, ringing, and chroma subsampling introduced by JPEG compression are ubiquitous in web-sourced imagery. Restormer achieves 33.19 dB on LIVE1 at quality factor 10, compared to 28.98 dB for ARCNN - a 4.21 dB improvement representing a fundamental advance in compression artifact suppression.

Normalized Multi-Metric Comparison

The chart below provides a normalized view across six performance dimensions simultaneously, enabling holistic comparison of architectures that excel in different respects.

Figure 3: Restoring Image Quality With AI using Real-ESRGAN and SwinIR

Key Cross-Cutting Findings

Hybrid architectures combining convolutional feature extraction with attention-based global reasoning consistently outperform purely convolutional or purely attention-based designs on most benchmarks.

The perceptual-distortion trade-off is empirically confirmed: GAN and diffusion models optimize different objective landscapes, and the choice must be guided by application requirements.

Real-world noise distributions remain substantially harder than synthetic counterparts.

Domain-adapted architectures outperform general-purpose models despite lower parameter counts.

Diffusion models represent the frontier of perceptual quality but require order-of-magnitude higher inference compute, limiting applicability in latency-sensitive deployments.

Self-supervised methods (Noise2Noise, Blind2Unblind) show strong promise for domains where paired training data is scarce or expensive to acquire.

VI. FUTURE SCOPE

Efficiency Attention for High-Resolution Processing

Processing native 4K or higher resolution images through attention-based architectures remains computationally prohibitive. Future research should investigate linear attention approximations, token pruning strategies, and hardware-

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 6, May 2026
Impact Factor: 8.2

aware architecture search. Mamba-based state-space models offer a theoretically promising O(N) complexity alternative and merit systematic evaluation on restoration benchmarks.

Physics-Informed Neural Restoration

Many imaging domains possess well-characterized degradation physics optical point spread functions in astronomy, X-ray attenuation models in CT imaging, or underwater light scattering. Integrating these physics constraints as differentiable layers within neural architectures could substantially reduce data requirements and improve generalization.

Federated and Privacy-Preserving Learning

Medical imaging restoration models require training on patient data, raising privacy concerns under GDPR and HIPAA. Federated learning protocols that train models across distributed hospital nodes without centralizing raw data represent an important future direction. Ensuring model quality under non-IID data distributions and communication constraints is an open challenge.

Multimodal and Cross-Sensor Fusion

Many real-world sensing systems deploy multiple modalities simultaneously: RGB cameras paired with depth sensors, thermal imagers, or LiDAR. Cross-modal restoration using high-quality modalities to guide degraded modality reconstruction is underexplored. Transformer architectures with cross-attention between modalities provide a natural framework.

Neuromorphic and On-Device Inference

As edge Al proliferates, restoration models must operate within milliwatt power budgets on embedded processors. Co-design of network architectures with target hardware (sparse inference engines, neuromorphic chips) and structured pruning, knowledge distillation, and quantization remain rich frontiers for enabling real-time on-device video restoration.

VII. LIMITATIONS

Benchmark Saturation and Metric Misalignment

Established benchmarks such as BSD68 and Set5 are approaching saturation, with top methods differing by fractions of a decibel. PSNR and SSIM correlate imperfectly with human perceptual quality; progress on these metrics does not guarantee meaningful improvement in application settings. The community requires new benchmarks with greater ecological validity.

Distribution Shift Between Training and Deployment

Deep learning restoration models are trained on specific noise and degradation distributions and often degrade significantly when applied to out-of-distribution conditions. A denoiser trained on SIDD smartphone images may perform poorly on scientific CCD noise or satellite imagery. Robustness to distribution shift remains fundamentally unsolved.

Hallucination and Fidelity Risks

GAN and diffusion-based methods occasionally synthesize plausible-looking but factually incorrect details - hallucinating textures or structures not present in the original scene. In high-stakes domains such as medical imaging or forensic analysis, such hallucinations carry significant harm potential. Uncertainty quantification in current methods is largely absent.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 6, May 2026 Impact Factor: 8.2

Computational and Environmental Cost

Training state-of-the-art restoration networks requires substantial GPU-hours and energy expenditure. Diffusion models trained for high-resolution restoration may consume thousands of A100-hours. The environmental footprint of this computation is rarely reported in publications and should be accounted for in assessments of method utility.

Interpretability and Explainability

Current deep restoration models operate largely as black boxes. It is generally unclear which training examples, architectural components, or feature representations drive output quality. Mechanistic interpretability remains an open problem for restoration architectures, limiting trust in safety-critical deployments.

VIII. CONCLUSION

This paper has presented a comprehensive examination of deep learning approaches to image restoration, noise reduction, and visual enhancement. Beginning from the theoretical formulation of image degradation as an ill-posed inverse problem, we traced the evolution of the field from early CNN baselines through GAN-based perceptual frameworks to contemporary transformer and diffusion-based architectures.

Our empirical analysis across four benchmark datasets confirms that transformer architectures particularly those employing efficient window-based or channel-wise attention - currently represent the best balance between quantitative performance and computational tractability for pixel-fidelity-oriented tasks. For perceptual quality, diffusion models achieve unmatched photorealism at substantial inference cost. Real-world noise distributions continue to pose distinctive challenges, rewarding domain-specific architectural adaptations.

Several fundamental tensions remain unresolved. The perceptual-distortion trade-off limits any single model from simultaneously optimizing for pixel accuracy and visual realism. Generalization across degradation types remains fragile. Computational costs of the most capable models present barriers to resource-constrained deployments.

Nevertheless, the trajectory of progress is striking. As research addresses the open challenges articulated here efficient attention, physics-informed priors, uncertainty quantification, and interpretability - deep learning is poised to become the definitive computational tool for image quality improvement across scientific, industrial, and everyday imaging applications.

REFERENCES

[1]. Zhang, K., Zuo, W., Chen, Y., Meng, D., & Zhang, L. (2017). Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising. IEEE Transactions on Image Processing, 26(7), 3142-3155.

[2]. Ledig, C., Theis, L., Huszar, F., Caballero, J., Cunningham, A., Acosta, A., & Shi, W. (2017). Photo-realistic single image super-resolution using a generative adversarial network. Proceedings of IEEE CVPR, 4681-4690.

[3]. Wang, X., Yu, K., Wu, S., Gu, J., Liu, Y., Dong, C., & Loy, C. C. (2018). ESRGAN: Enhanced super-resolution generative adversarial networks. Proceedings of ECCV Workshops, 63-79.

[4]. Liang, J., Cao, J., Sun, G., Zhang, K., Gool, L. V., & Timofte, R. (2021). SwinIR: Image restoration using Swin Transformer. Proceedings of IEEE ICCV Workshops, 1833-1844.

[5]. Zamir, S. W., Arora, A., Khan, S., Hayat, M., Khan, F. S., & Yang, M. H. (2022). Restormer: Efficient transformer for high-resolution image restoration. Proceedings of IEEE CVPR, 5728-5739.

[6]. Ho, J., Jain, A., & Abbeel, P. (2020). Denoising diffusion probabilistic models. Advances in NeurIPS, 33, 6840-6851.

[7]. Chen, L., Chu, X., Zhang, X., & Sun, J. (2022). Simple baselines for image restoration (NAFNet). Proceedings of ECCV, 17-33.

[8]. Dosovitskiy, A., Beyer, L., Kolesnikov, A., et al. (2021). An image is worth 16x16 words: Transformers for image recognition at scale. Proceedings of ICLR.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 6, May 2026
Impact Factor: 8.2

[9]. Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. Proceedings of MICCAI, 234-241.

[10]. Zhang, Y., Li, K., Li, K., Wang, L., Zhong, B., & Fu, Y. (2018). Image super-resolution using very deep residual channel attention networks. Proceedings of ECCV, 286-301.

[11]. Lehtinen, J., Munkberg, J., Hasselgren, J., et al. (2018). Noise2Noise: Learning image restoration without clean data. Proceedings of ICML, 2965-2974.

[12]. Abdelhamed, A., Lin, S., & Brown, M. S. (2018). A high-quality denoising dataset for smartphone cameras. Proceedings of IEEE CVPR, 1692-1700.

[13]. Johnson, J., Alahi, A., & Fei-Fei, L. (2016). Perceptual losses for real-time style transfer and super-resolution. Proceedings of ECCV, 694-711.

[14]. Restoring Image Quality With AI using Real-ESRGAN and SwinIR | by Maria Llain | Medium