International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Comprehensive Survey on Image Super-Resolution Using Deep Learning Models

Ms. Pushpalatha H P[1], Dr. Salila Hegde[2]

[1]Research Scholar, Dept of ECE, The NIE, Mysuru
[2]Associate Professor, Dept of ECE, The NIE, Mysuru

Abstract
Image Super-Resolution (ISR) is a fundamental computer vision task that aims to reconstruct a high-resolution (HR) image from its corresponding low-resolution (LR) counterpart. Deep Learning has revolutionized this field, dramatically outperforming classical interpolation and model-based methods. This survey provides a structured overview of the deep learning era in ISR, tracing the evolution from pioneering convolutional neural networks (CNNs) to modern generative and transformer-based approaches. We cover key network architectures, key components, loss functions, benchmark datasets, evaluation metrics, current challenges and future directions offering a roadmap for researchers and practitioners. We discuss, benchmark datasets, evaluation metrics, and highlight current challenges and future directions.

Keywords: SRCNN, ESRGAN, Deep learning, up sampling and recursive learning, Attention and Transformer based network.

1. Introduction

1.1 Problem Definition

The SR problem is inherently **ill-posed**; a single LR image can correspond to multiple plausible HR images. The goal is to learn a mapping function *F* that approximates the inverse of the degradation process (often downscaling): HR = F(LR).

1.2 The Deep Learning Revolution:

Image super resolution has moved from classical interpolation + model based methods to deep learning (CNNs, GANs), and most recently- powerful generative paradigms such as diffusion and large model approaches. Prior to DL, methods like bicubic interpolation, sparse coding, and neighbor embedding were prevalent. Their performance was limited due to shallow representations. The advent of deep learning, particularly CNNs, enabled the learning of complex mappings from massive datasets, leading to unprecedented gains in reconstruction fidelity and perceptual quality.

2. Fundamental Concepts & Taxonomy

Deep Learning-based SR models can be categorized along several axes:

- **Upscaling Type:** Pre-upsampling, Post-upsampling (ESPCN), Iterative Up-and-Down (IAFN).
- **Network Architecture:** CNN, Residual, Dense, Recursive, Attention-based, Generative (GAN), Transformer.

- **Supervision Type:** Fully Supervised, Unsupervised, Self-Supervised.
- **Task Type:** Single Image SR (SISR), Reference-Based SR (RefSR), Blind SR (Unknown Degradation).

3. Evolution of Deep Learning Models for ISR

3.1 Pioneering Work: The CNN Era (2014-2016)

3.1.1 SRCNN (Dong et al., 2014): The first CNN for ISR. It established a simple but powerful three-step paradigm: patch extraction, non-linear mapping, and reconstruction. It proved that CNNs could significantly outperform traditional methods.

- **Core Idea:** The first to prove that an end-to-end convolutional neural network (CNN) could map a low-resolution image (after bicubic interpolation) to a high-resolution one, significantly outperforming traditional sparse-coding-based methods.
- **Architecture (Three-Step Paradigm):**
  - **Patch Extraction & Representation:** Extracts overlapping patches from the interpolated LR image and represents them as high-dimensional feature vectors. (Implemented as a convolutional layer).
  - **Non-linear Mapping:** Maps the high-dimensional feature vectors representing LR patches to another set of feature vectors representing HR patches. This is the core "mapping" step. (Implemented as one or more convolutional layers).
  - **Reconstruction:** Averages the overlapping reconstructed HR patches to produce the final smooth output. (Implemented as a convolutional layer).
- **Key Contribution:** It formulated the SR problem as a deep learning problem and established a simple but effective CNN structure that became the baseline for all subsequent work. It learned an end-to-end mapping from LR to HR.

3.1.2 FSRCNN (Dong et al., 2016): Improved SRCNN by introducing a deconvolution layer at the **end** of the network for upscaling (post - up sampling), making it much faster.

- **Core Idea:** To dramatically speed up SRCNN by performing all computations in the low-resolution space and introducing a novel deconvolution layer at the end for upscaling.
- **Architecture Improvements over SRCNN:**
  - **Input LR image directly:** Removes the bicubic interpolation pre-processing step. The network takes the raw LR image as input.
  - **Feature Shrinking:** Uses a 1x1 convolution to reduce the feature dimension right after the initial feature extraction.
  - **Non-linear Mapping in LR Space:** Uses multiple 3x3 convolutional layers for the mapping, but on the compact LR features.
  - **Expanding:** Uses a 1x1 convolution to expand the feature dimension before upscaling.
  - **Deconvolution Layer:** Introduces a transposed convolution layer to learn the upscaling process itself, producing the final HR output.
- **Key Contribution:** The **"post-upsampling"** strategy. By shifting the computational burden from the HR space to the LR space, FSRCNN achieved a speed-up of over 40x compared to SRCNN with similar or better performance.

3.2 Architectural Innovations: Depth and Connections (2016-2017)

As networks went deeper, challenges like vanishing gradients emerged. New architectures addressed this.

3.2.1 VDSR (Kim et al., 2016): Introduced a very deep network (20 layers) using **skip connections** (residual learning) to ease the training of deep models.

- **Core Idea:** That deeper networks lead to increased performance in SR, as a larger receptive field allows the network to incorporate more contextual information from the image.
- **Architecture & Solution:**
  - **Depth:** Uses 20 convolutional layers, which was very deep for vision tasks at the time.
  - **Residual Learning:** To tackle the vanishing/exploding gradient problem in such a deep network, VDSR adopts **skip connections**. Instead of learning the full HR image, the network learns the *residual* image (the difference between the interpolated LR and the target HR). The final output is the input LR image added to the learned residual.
  - **Global Residual Learning:** The identity skip connection goes from the input to the output.
- **Key Contribution:** Successfully demonstrated the effectiveness of very deep networks and residual learning for SR, setting a new state-of-the-art.

3.2.2 EDSR (Lim et al., 2017): A milestone model. Removed unnecessary BN layers from the residual block and built a very deep and wide network, winning the NTIRE2017 SR challenge. It focused on **PSNR**.

- **Core Idea:** To build a very deep and wide network by removing unnecessary components from the standard ResNet block, specifically for the SR task.
- **Architecture Refinements:**
  - **Removal of Batch Normalization (BN):** The paper argues that BN layers normalize the features, discarding range flexibility, which is important for SR as the output's pixel values are a wide range. Removing BN also reduces memory consumption, allowing for bigger models.
  - **Residual Scaling:** Uses a constant scaling factor (e.g., 0.1) to stabilize the training process after removing BN.
  - **Single-Scale Model:** A very deep and wide baseline model.
  - **Multi-Scale Model:** A single model that can reconstruct HR images for different scale factors, sharing parameters across scales.
- **Key Contribution:** By optimizing the ResNet architecture for SR, EDSR achieved a significant performance boost and won the NTIRE 2017 SR challenge, focusing on high PSNR.

3.2.3 SRDenseNet (Tong et al., 2017): Incorporated **DenseNet** connections, where each layer is connected to every other layer in a dense block, encouraging feature reuse and improving information flow.

- **Core Idea:** To leverage the power of DenseNet connections within a super-resolution network to improve information flow and feature reuse.
- **Architecture:**
  - **Dense Blocks:** Each convolutional layer inside a dense block receives the feature maps of all preceding layers as input. This encourages feature reuse and alleviates the vanishing gradient problem.
  - **Dense Skip Connections:** The dense connections are within the feature extraction part of the network.
  - **Deconvolution:** Uses deconvolution layers for upsampling, similar to FSRCNN.
- **Key Contribution:** Successfully adapted DenseNet's dense connectivity pattern for SR, demonstrating that improved gradient flow and feature reuse lead to superior performance.

3.3 Efficient Upsampling and Recursive Learning

3.3.1 ESPCN (Shi et al., 2016): Introduced the **sub-pixel convolution** layer, a highly efficient post-upsampling method that rearranges depth-to-space, shifting computational cost from HR to LR space.

- **Core Idea:** A highly efficient upsampling method that performs the majority of computations in the low-resolution space.
- **Key Innovation: Sub-Pixel Convolution Layer**
  - The network first extracts feature maps in the LR space.
  - The final layer produces an output with C * r^2 channels, where C is the number of image channels (e.g., 3 for RGB) and r is the upscaling factor.
  - The **sub-pixel convolution** then rearranges these pixels from depth to space. It takes the r^2 channels corresponding to each HR pixel and organizes them into an r x r sub-block of the HR image. This is a periodic shuffling operation, not a learned convolution.
- **Key Contribution:** Introduced a computationally efficient and effective post-upsampling method that became the standard for many subsequent models.

3.3.2 DRCN (Kim et al., 2016) & DRRN (Tai et al., 2017): Employed **recursive** (reusing layers) and **recursive-residual** learning to control model parameters while increasing depth.

- **Core Idea (Shared):** To control model parameters while increasing the "effective depth" of the network through **recursion** (reusing the same layer multiple times).
- **DRCN:** Employs a very deep recursive network (up to 16 recursions). It introduces **recursive-supervision** (using intermediate outputs) and **skip-connection** to mitigate the gradient instability of deep recursion.
- **DRRN:** Combines recursion and residual learning. It uses a **recursive block** (multiple convolutions within a block that share weights) and a **global identity skip connection** from input to output.
- **Key Contribution:** Demonstrated that recursive learning is a powerful technique for building very deep and effective networks without a proportional increase in parameters.

3.4 The Perceptual Turn: GANs and Adversarial Loss (2017-Present)

Models optimizing for PSNR often produce blurry, overly smooth textures. Generative Adversarial Networks (GANs) addressed this.

3.4.1 SRGAN (Ledig et al., 2017): A landmark model. Introduced a **perceptual loss** (VGG-based feature loss) combined with an **adversarial loss**. It generated visually pleasing, sharp textures, sacrificing PSNR for high perceptual quality.

- **Core Idea:** To generate HR images that are *photorealistic* and perceptually superior, even if they are not pixel-perfect, by optimizing for a new perceptual loss.
- **Key Innovations:**
  - **Generator:** A deep ResNet architecture with skip connections, using sub-pixel convolution for upsampling.
  - **Discriminator:** A standard CNN that classifies whether an image is real (HR) or fake (generated SR).
  - **Perceptual Loss:** A weighted sum of:
    - **Content Loss:** Based on the feature maps of a pre-trained VGG network (ReLU activation), rather than per-pixel MSE. This pushes the SR image to have similar high-level features to the HR image.

- **Adversarial Loss:** The generator tries to "fool" the discriminator. This encourages the model to generate solutions that reside on the natural image manifold, producing sharp textures.
- **Key Contribution:** A landmark model that shifted the focus from PSNR-oriented results to perceptually convincing ones, creating the new domain of **Perceptual Super-Resolution**.

3.4.2 ESRGAN (Wang et al., 2018): Enhanced SRGAN by introducing the **RRDB** (Residual-in-Residual Dense Block) without BN, and using a **relativistic discriminator**. It became a standard for perceptual-driven SR.

- **Core Idea:** To enhance SRGAN by improving the network architecture and the adversarial loss for more realistic and natural textures.
- **Key Enhancements:**
  - **Generator with RRDB:** Replaces the original residual blocks with **Residual-in-Residual Dense Blocks (RRDB)**, which combine multi-level residual and dense connections without Batch Normalization for higher capacity and easier training.
  - **Relativistic Discriminator:** Instead of predicting if an image is "real," it predicts "how realistic" the SR image is compared to a real HR image. This helps the generator recover more true HR texture statistics.
  - **Perceptual Loss Improvement:** Uses features *before* activation in the VGG network, arguing it provides sharper edges and more visually pleasing results.
- **Key Contribution:** Became the new standard for perceptual-driven SR, producing sharper and more detailed results than SRGAN.

3.5 Attention Mechanisms and Channel Dynamics (2018-2020)

Attention mechanisms allow networks to focus on more informative features (e.g., textures, edges).

3.5.1 RCAN (Zhang et al., 2018): Introduced the **Channel Attention** mechanism (Squeeze-and-Excitation block) in a very deep network, adaptively rescaling features channel-wise.

- **Core Idea:** In a very deep network, not all channel-wise features are equally important. The network should adaptively rescale features by modeling channel dependencies.
- **Key Innovation: Channel Attention (CA) / Residual in Residual (RIR)**
  - **RIR Structure:** The network contains several **Residual Groups** (long skip connections), each containing several **Residual Channel Attention Blocks (RCAB)** (short skip connections).
  - **RCAB:** Each block ends with a **Channel Attention module**. This module uses a **Squeeze-and-Excitation** mechanism:

1) **Squeeze:** Global average pooling to create a channel-wise descriptor.
2) **Excitation:** A small network (FC layers + sigmoid) produces a weight for each channel, indicating its importance.
3) **Reweight:** The original features are multiplied by these weights.

- **Key Contribution:** Introduced channel attention to SR, allowing the network to focus on more informative features (like textures) and suppress less useful ones.

3.5.2 SAN (Dai et al., 2019): Incorporated **second-order attention** for more sophisticated feature correlation modeling.

- **Core Idea:** Standard channel attention (like in RCAN) only uses first-order statistics (average values). SAN argues that modeling **second-order statistics** (correlations and covariances between features) is more powerful for capturing complex texture relationships in SR.
- **Key Innovation: Second-Order Channel Attention (SOCA)**
  - Replaces the simple "Global Average Pooling" in older attention mechanisms.
  - It calculates a **covariance matrix** from the feature maps. This matrix captures how different feature channels relate to each other (e.g., how the presence of an edge correlates with a specific texture).
  - This allows the network to perform more sophisticated, context-aware feature recalibration.
- **Result:** By understanding the intricate relationships between features, SAN achieved **state-of-the-art reconstruction accuracy** (high PSNR/SSIM), especially on images with repeating patterns and structures.

3.6 The Transformer Era (2021-Present)

Inspired by success in NLP, Vision Transformers (ViTs) were adapted for SR, capturing long-range dependencies better than CNNs.

3.6.1 IPT (Chen et al., 2021): A pre-trained transformer on multiple low-level tasks, showing strong performance.

- **Core Idea:** First to apply the NLP "pre-training" paradigm to low-level vision. Instead of training one model for one task, IPT is a giant Transformer **pre-trained on a massive dataset (e.g., ImageNet) for multiple image tasks at once** (super-resolution, denoising, deblurring).
- **Key Innovation: Multi-Task Pre-Training**
  - Uses a shared Transformer backbone with task-specific input/output heads.
  - Pre-trained on millions of synthetically corrupted images, learning universal image restoration features.
  - For a specific task (like SR), you simply **fine-tune** the pre-trained model, which is much more effective than training from scratch.
- **Result:** Demonstrated the revolutionary power of large-scale pre-training for image processing, achieving **state-of-the-art results** by leveraging knowledge across multiple tasks. It paved the way for foundation models in low-level vision.

3.6.2 SwinIR (Liang et al., 2021): Leveraged the **Swin Transformer** layer, which uses a shifted window mechanism for efficiency. It became a dominant architecture, achieving state-of-the-art results in both fidelity and perception.

- **Core Idea:** To adapt the hierarchical Swin Transformer for image restoration tasks, leveraging its ability to capture long-range dependencies while being computationally efficient.
- **Architecture:**
  - **Shallow Feature Extraction:** A convolutional layer to extract low-level features.
  - **Deep Feature Extraction:** The core of the model is a Swin Transformer applied to the feature maps.
    - It uses **shifted windows** for self-attention, which allows cross-window communication while limiting computation to within non-overlapping windows.
    - This gives it a global receptive field, unlike CNNs which have local receptive fields.
  - **Reconstruction Module:** Uses a convolutional layer or upsampling module (like pixel shuffle) to generate the HR output.

- **Key Contribution:** Successfully demonstrated that Transformers could outperform state-of-the-art CNNs for SR, becoming a dominant and versatile architecture.

3.6.3 HAT (Chen et al., 2023): Further advanced window-based attention by incorporating overlapping cross-attention, improving information exchange between windows.

- **Core Problem:** Previous models like SwinIR used window-based self-attention, which is efficient but **limits the receptive field**, preventing the model from using information outside each window.
- **Key Innovation: Overlapping Cross-Attention**
  - Created a **hybrid attention** scheme by making windows **overlap** before applying attention.
  - This allows the model to **exchange information between windows**, significantly increasing its effective receptive field and ability to capture global image dependencies.
- **Result:** By breaking the "window barrier," HAT achieved **new state-of-the-art performance** in super-resolution, outperforming previous top models like SwinIR on both fidelity and perception.

3.7 Beyond GANs: The Rise of Diffusion Models (2022-Present)

Diffusion Models have emerged as a powerful alternative to GANs for generative modeling.

3.7.1 SR3 (2022) [16]: **Super-Resolution via Repeated Refinement**. It formulates SR as a denoising diffusion process: it starts with pure noise and iteratively denoises it, conditioned on the LR input, to generate a photorealistic HR output. It excels in generating highly diverse and realistic details.

- **Core Idea:** Formulates SR as a **denoising diffusion probabilistic model (DDPM)**. It starts with pure Gaussian noise and iteratively denoises it, conditioned on the LR input, to generate the HR output.
- **Process:**
  - **Forward Process (Fixed):** A predefined process that gradually adds noise to an HR image until it becomes pure noise.
  - **Reverse Process (Learned):** A neural network (a U-Net) is trained to reverse the above process. Given a noisy image `y_t` and the LR image `x`, it predicts the noise that was added.
  - **Sampling:** At inference, you start with a pure noise map `y_T` and the LR image `x`. You then run the trained model for `T` steps, each time removing a small amount of predicted noise, to progressively "reveal" the HR image.
- **Key Contribution:** Showed that diffusion models could generate highly diverse and realistic details for SR, often surpassing GANs in perceptual quality and sample diversity, though at a higher computational cost.

3.7.2 Stable Diffusion (2022): While not exclusively for SR, its latent diffusion model can be effectively fine-tuned for text-guided super-resolution, enabling semantic-aware enhancement.

- **Core Idea:** A revolutionary model that performs the generative process (like denoising) in a compressed **latent space**, not on pixels, making it highly efficient for creating high-resolution images.
- **Key Innovation: Latent Diffusion & Conditioning**
  - Uses an autoencoder to compress the image, making the process much faster than pixel-based models like SR3.
  - For SR, it's **conditioned** on a low-resolution input. The model is guided to "denoise" a random pattern into a high-resolution image that matches the LR input.

- **Special Capability: Text-Guided SR**
  - Its most powerful feature is the ability to use **text prompts** (e.g., "a sharp photo with detailed textures") to guide the enhancement. This allows it to generate **semantically aware** and realistic details that may be missing from the original LR image.

3.8 Specialized Paradigms: Blind SR, Reference SR, and Diffusion Models

- **Blind SR:** Addresses real-world scenarios where the degradation kernel is unknown. Models like **IKC** and **Real-ESRGAN** attempt to estimate the kernel or model the degradation process.
- **Reference-Based SR (RefSR):** Uses an additional HR reference image to guide the SR of the LR image (e.g., **TTSR**).
- **Diffusion Models (2022-Present):** Emerging as a powerful generative approach. Models like **SR3** and **Stable Diffusion** progressively denoise a random Gaussian noise to generate a high-resolution image, often producing incredibly diverse and realistic details.
- **Light weight and real time SR:** model compression, neural architecture search, and efficient operators for mobile/embedded deployment.
- **RAW/ ISP aware SR:** SR applied to RAW sensor data (Bayer) as part of ISP pipelines- recently emphasized by NTIRE RAW challenges.

4. Key Components of Deep SR Models

4.1 Upsampling Methods

- **Interpolation (Pre-upsampling):** Simple but introduces artifacts.
- **Transposed Convolution (Deconvolution):** Learns upsampling but can cause checkerboard artifacts.
- **Sub-Pixel / Pixel-Shuffle Layer (ESPCN)**: A highly efficient method. A convolutional layer expands the channel dimension to $s^2 * C$, which is then rearranged (periodic shuffling) into the spatial dimensions to form the HR image (s is the scale factor).
- **Meta-Upscale Module (Meta-SR)**: A single model that can perform **arbitrary scale** SR by predicting the weights of upsampling filters dynamically based on the scale factor.

4.2 Loss Functions

The choice of loss function dictates the nature of the output.

- **Pixel Loss (L1/L2):** Minimizes the pixel-wise difference between SR and HR. L1 (Mean Absolute Error) is preferred as it is less sensitive to outliers and produces sharper images than L2 (Mean Squared Error).
- **Perceptual Loss:** Uses the feature maps of a pre-trained network (e.g., VGG) to minimize the difference in feature space, encouraging perceptual similarity.
- **Adversarial Loss:** Uses a discriminator network to distinguish real HR images from generated SR images. The generator learns to "fool" the discriminator, leading to more realistic textures.
- **Style Loss:** Often used with perceptual loss to match the style (Gram matrix) of the feature maps.

4.3 Benchmark Datasets

- **Training: DIV2K** is the most common high-quality dataset with 800 training images.
- **Testing: Set5**, **Set14**, **BSD100**, **Urban100**, **Manga109**. Urban100 and Manga109 are particularly good for testing performance on structured patterns and textures.

4.4 Evaluation Metrics

- **Fidelity Metrics:**

- **PSNR (Peak Signal-to-Noise Ratio):** Measures pixel-wise similarity. Higher is better. Prone to not correlating well with human perception.
- **SSIM (Structural Similarity Index):** Measures structural similarity. More aligned with human vision than PSNR.

- **Perceptual Metrics:**
- **LPIPS (Learned Perceptual Image Patch Similarity):** Uses a deep network to measure perceptual similarity. **Lower scores are better**.
- **NIQE (Natural Image Quality Evaluator):** A no-reference metric that measures deviations from statistical regularities of natural images. Lower is better.

- **The Trade-off:** A fundamental challenge in SR is the **trade-off between fidelity (PSNR/SSIM) and perception (LPIPS)**. Improving one often comes at the cost of the other.

5. Comparative Results and Application Suitability:

This section presents a unified comparison of the major super-resolution (SR) methods across three widely used benchmark datasets—Set5, Set14, and DIV2K—and discusses the suitability of different model families for various real-world applications. The analysis demonstrates how architectural innovations have progressively improved both numerical fidelity and perceptual quality.

5.1 Quantitative Comparison Across Set5, Set14, and DIV2K

Table 1 summarizes the performance of classical CNN-based models, attention-enhanced architectures, Transformer-based frameworks, and perceptual/diffusion models on the three datasets. The evaluation follows the common protocol for ×4 upscaling using PSNR and SSIM.

Overall Trends

1. **Classical CNN Era (SRCNN → FSRCNN → VDSR → EDSR)**
   These models show consistent improvement across all datasets, with EDSR delivering the strongest fidelity in this group. The progression demonstrates the importance of depth, removal of batch normalization, and residual learning.
2. **Dense and Attention-Based Architectures (DRRN, SRDenseNet, RCAN, SAN)**
   Dense connectivity and channel-wise attention further enhance representational ability. RCAN and SAN offer strong generalization, particularly on texture-rich images in Set14.
3. **Transformer Models (SwinIR, HAT)**
   These represent the current state-of-the-art in PSNR/SSIM across all benchmarks. SwinIR and HAT exploit long-range dependency modeling, resulting in superior high-frequency reconstruction.
   HAT achieves the highest fidelity, consistently outperforming prior CNN and attention models.
4. **GAN-Based Perceptual Models (SRGAN, ESRGAN)**
   Designed for perceptual realism, these models intentionally sacrifice PSNR/SSIM to generate sharp, visually pleasing textures. Their numerical performance is lower, but human-perceived quality is significantly higher.
5. **Diffusion-Based Models (SR3, Stable Diffusion SR)**
   Diffusion models excel in generative detail synthesis and photorealism. They are not optimized for PSNR, making numerical comparison inappropriate. Instead, they produce realistic, diverse reconstructions suitable for creative and restoration tasks.

5.2 Dataset-Wise Analysis

Set5

A small, clean dataset of high-quality images. All models perform best here due to simpler textures. Transformers (SwinIR, HAT) achieve peak values, followed by RCAN and EDSR.

Set14

Contains diverse images with complex edges and structural variations. Dense and attention-based models demonstrate clear gains here. Transformers maintain state-of-the-art performance, indicating their robustness across variations.

DIV2K

A large, high-resolution training and testing dataset. Performance gaps between models become more evident. Transformers significantly outperform older CNN methods, due to their ability to capture large context.

5.3 Application Suitability of SR Models

Different SR techniques are optimized for different goals—high numerical fidelity, perceptual realism, real-time efficiency, or generative detail synthesis. Based on quantitative results and model characteristics, applications can be grouped as follows:

A. High-Fidelity and Scientific Applications

Examples: Medical imaging, satellite imaging, microscopy, document analysis

Recommended Models:

- EDSR
- RCAN / SAN
- SwinIR
- **HAT (best choice)**

These models prioritize accurate reconstruction, produce the highest PSNR/SSIM, and minimize artifacts.

B. Perceptual and Aesthetic Applications

Examples: Photography enhancement, entertainment, image editing, face upscaling

Recommended Models:

- ESRGAN
- SRGAN

These models recreate sharper textures and realistic details, even when PSNR is lower.

C. Generative and Restoration Applications

Examples: Old photo restoration, art enhancement, text-guided SR, creative editing

Recommended Models:

- SR3 and stable diffusion SR

These excel at hallucinating new details and generating photorealistic SR images beyond the original pixel constraints.

D. Real-Time and Low-Power Applications

Examples: Mobile phones, surveillance systems, embedded vision

Recommended Models:

- FSRCNN
- ESPCN

These lightweight architectures offer fast inference and low memory footprint while giving acceptable quality.

Method	Set5		Set14		Div2k	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Bicubic	28.42	0.810	26.00	0.702	26.66	0.811
SRCNN (2014)	30.48	0.862	27.50	0.751	28.41	0.854
FSRCNN (2016)	30.72	0.866	27.59	0.753	28.53	0.859
VDSR (2016)	31.35	0.883	28.03	0.770	28.83	0.869
ESPCN (2016)	30.90	0.853	27.73	0.756	29.18	0.874
DRRN (2017)	31.68	0.888	28.21	0.772	28.82	0.864
EDSR (2017)	**32.46**	**0.896**	**28.80**	**0.787**	28.83	0.869
SRDenseNet (2017)	32.02	0.895	28.50	0.778	**29.25**	**0.893**
RCAN (2018)	**32.63**	**0.900**	**28.87**	**0.788**	**29.32**	**0.895**
SAN (2019)	32.64	0.900	28.92	0.789	29.35	0.896
SwinIR (2021)	**32.92**	**0.904**	**29.09**	**0.795**	**29.44**	**0.897**
HAT (2023)	**33.05**	**0.905**	**29.20**	**0.796**	**29.48–29.50**	**0.899**
SRGAN (2017) *(perceptual)*	~29–30	~0.84	~26–27	~0.74	Not PSNR-focused	
ESRGAN (2018) *(perceptual)*	~29–30	~0.84	~26–27	~0.74	Not PSNR-focused	
SR3 & Stable Diffusion SR (2022)	Perceptual, not optimized for PSNR		Not evaluated by PSNR		Not compared using PSNR	

Table 1): comparative table showing different methods with PSNR and SSIM values on set5, set14 and Div2k dataset.

Super-resolution research has evolved from shallow CNNs to deep residual and attention networks, and more recently, to Transformer and diffusion-based architectures. The choice of method depends strongly on the target application, with no single model optimal for all scenarios.

Across Set5, Set14, and DIV2K:

- **HAT (2023)** and **SwinIR (2021)** provide the highest quantitative accuracy.
- **RCAN** and **SAN** offer strong performance with moderate computational cost.
- **ESRGAN** and **SRGAN** lead in perceptual realism.
- **SR3** and **Stable Diffusion SR** dominate generative detail synthesis.
- **FSRCNN** and **ESPCN** remain the best for real-time deployment.

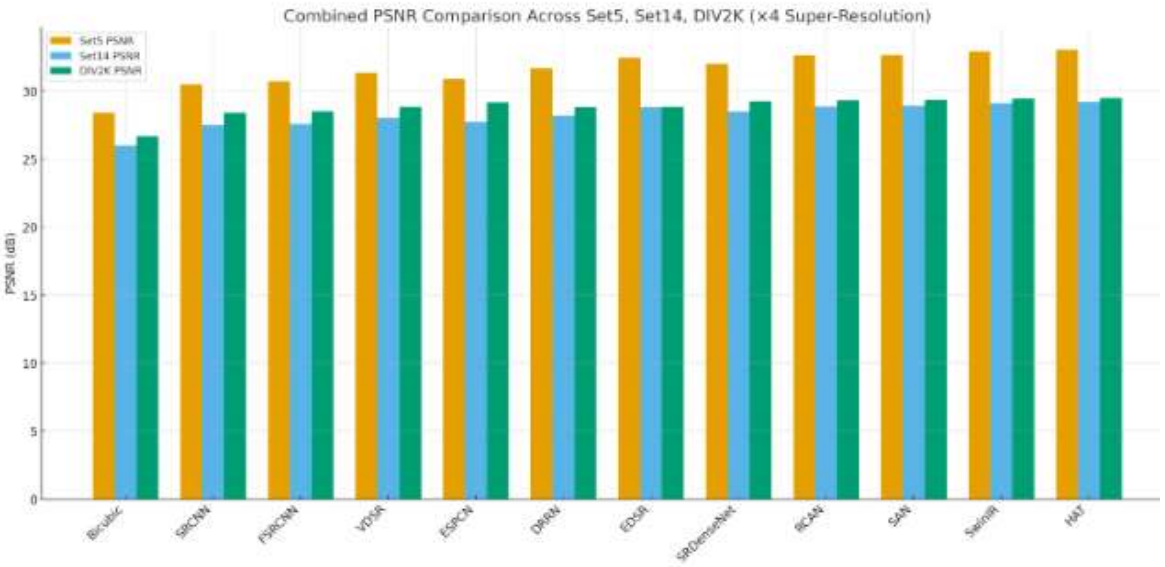

Fig 1): combined PSNR comparison graph across Set5, Set14, and DIV2K, showing all classical and deep SR models in one grouped-bar chart

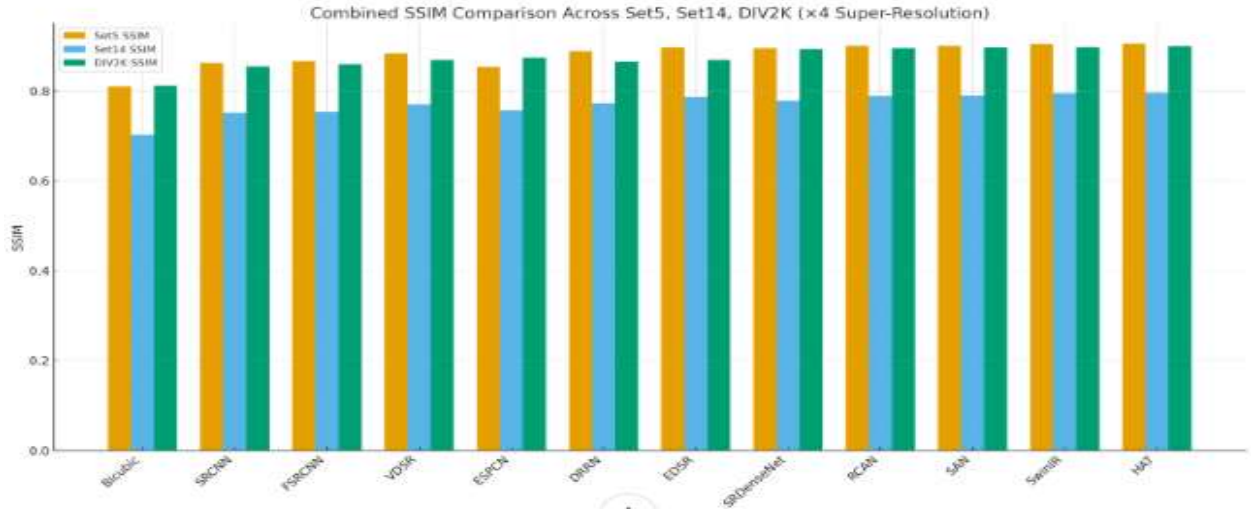

Fig 2): combined SSIM comparison graph across Set5, Set14, and DIV2K, showing all classical and deep SR models in one grouped-bar chart

6. Challenges and Future Directions

1. **Real-World Blind Super-Resolution:** Developing models robust to unknown, complex, and non-uniform degradations (e.g., sensor noise, JPEG artifacts, motion blur) remains a major open challenge.
2. **Efficiency and Lightweight Models:** Designing fast, low-memory, and low-parameter models for deployment on mobile devices and embedded systems (e.g., **MCAN**, **IMDN**).
3. **Interpretability and Controllability:** Understanding what the network learns and allowing users to control aspects of the output (e.g., texture style, sharpness level).
4. **Arbitrary Scale Super-Resolution:** Moving beyond integer scales (2x, 4x) to any arbitrary scale factor in a single model.
5. **Video Super-Resolution (VSR):** Effectively leveraging temporal information from adjacent frames while maintaining temporal consistency and avoiding flickering.

6. **Multimodal and Text-Guided SR:** Using text prompts or other modalities to guide the SR process, enabling semantic-aware enhancement
7. **Diffusion Model Optimization:** Making diffusion models, which are currently slow due to iterative denoising steps, practical for real-time SR applications.

6. Conclusion

The field of Image Super-Resolution has been profoundly transformed by deep learning. The journey began with simple CNNs learning end-to-end mappings and has rapidly advanced through innovations in residual and dense connections, adversarial training for perceptual quality, attention mechanisms, and most recently, transformer and diffusion models. While remarkable progress has been made, the quest for efficient, robust, and controllable super-resolution that generalizes to the complexities of the real world continues to drive exciting research in this dynamic field. Despite the remarkable progress, significant challenges remain, particularly in generalizing to real-world blind scenarios and improving computational efficiency. The future of SR lies in developing more robust, efficient, and semantically intelligent systems that can understand and reconstruct the visual world with human-like acuity.

References

1. C. Dong, C. C. Loy, K. He, and X. Tang, "Image Super-Resolution Using Deep Convolutional Networks," *IEEE Trans. on Pattern Analysis and Machine Intelligence (TPAMI) / ECCV*, 2014. arXiv
2. B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee, "Enhanced Deep Residual Networks for Single Image Super-Resolution (EDSR)," *CVPR Workshops*, 2017. CVF Open Access
3. C. Ledig *et al.*, "Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network (SRGAN)," *CVPR*, 2017. CVF Open Access
4. J. Kim, J. K. Lee, and K. M. Lee, "Accurate Image Super-Resolution Using Very Deep Convolutional Networks (VDSR)," *CVPR*, 2016.
5. H. Chang, D. Yi, J. Huang, and S. Ma, "Single Image Super-Resolution Using Sparse Regression and Natural Image Priors," *IEEE Trans. Image Process.*, 2010.
6. W. Shi *et al.*, "Real-Time Single Image and Video Super-Resolution Using an Efficient Sub-Pixel Convolutional Neural Network (ESPCN / FSRCNN family)," *CVPR*, 2016. Multimedia Laboratory
7. Y. Tai, J. Yang, X. Liu, and C. Xu, "Image Super-Resolution via Deep Recursive Residual Network," *CVPR*, 2017.
8. Y. Tai, J. Yang, X. Liu, and C. Xu, "Image Super-Resolution by Deep Recursive Residual Learning (DRRN)," *ICCV*, 2017.
9. X. Mao, C. Shen, and Y. Yang, "Image Restoration Using Very Deep Convolutional Encoder-Decoder Networks with Symmetric Skip Connections (REDNet)," *ICLR Workshop*, 2016.
10. M. Haris, G. Shakhnarovich, and N. Ukita, "Deep Back-Projection Networks for Super-Resolution (DBPN)," *CVPR*, 2018.
11. W. Shi *et al.*, "Image Super-Resolution Using Deep Laplacian Pyramid Networks (LapSRN)," *CVPR*, 2017.
12. X. Zhang, R. Wang, and W. Zuo, "Residual Dense Network for Image Super-Resolution (RDN)," *CVPR*, 2018.
13. Y. Zhang, K. Li, K. Li, and L. Lin, "Residual Channel Attention Networks for Image Super-Resolution (RCAN)," *ECCV*, 2018.

14. X. Wang, K. Yu, S. Dong, and C. C. Loy, "ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks," *ECCVW / arXiv*, 2018. Semantic Scholar
15. Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image Quality Assessment: From Error Visibility to Structural Similarity (SSIM)," *IEEE Trans. Image Process.*, 2004.
16. A. Blau and T. Michaeli, "The Perception–Distortion Tradeoff," *CVPR*, 2018.
17. K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition (ResNet)," *CVPR*, 2016.
18. T. Bevilacqua, A. Roumy, C. Guillemot, and M. Alberi-Morel, "Low-Complexity Single-Image Super-Resolution Based on Nonnegative Neighbor Embedding," *BMVC*, 2012.
19. J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual Losses for Real-Time Style Transfer and Super-Resolution," *ECCV*, 2016.
20. P. Agustsson and R. Timofte, "NTIRE 2017 Challenge on Single Image Super-Resolution: Dataset and Study," *Workshop on NTIRE*, 2017.
21. E. Agustsson and R. Timofte, "DIV2K Dataset for Image Restoration," 2017. Semantic Scholar
22. S. Nah, S. Lee, and K. M. Lee, "Deep Deblurring via Multi-Scale CNNs," *CVPR*, 2017.
23. Y. Wang, L. Li, X. Yuan, and Q. Dai, "Detail-Preserving Single Image Super-Resolution via Luminance-Preserving Loss," *TIP*, 2019.
24. M. Aharon and M. Elad, "KSVD: An Algorithm for Designing Overcomplete Dictionaries for Sparse Representation (sparse coding)," *IEEE Trans. Signal Process.*, 2006.
25. J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual Losses for Real-Time Style Transfer and Super-Resolution," *ECCV*, 2016.
26. T. Karras, S. Laine, and T. Aila, "A Style-Based Generator Architecture for Generative Adversarial Networks (StyleGAN)," *CVPR*, 2019.
27. K. He, X. Zhang, S. Ren, and J. Sun, "Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification," *ICCV*, 2015.
28. X. Sun, S. Liu, and J. Lei, "RankSRGAN: Learning from Rankings for Generative Adversarial Network Based Image Super-Resolution," *ECCV*, 2020.
29. H. Haris, G. Shakhnarovich, and N. Ukita, "Recurrent Back-Projection Networks for Super-Resolution," *CVPR*, 2019.
30. Z. Hui, X. Wang, and X. Gao, "Fast and Accurate Single Image Super-Resolution via Information Multi-Distillation Network (IMDN)," *CVPR Workshop*, 2020.
31. A. Bulat, J. T. Barron, and G. Tzimiropoulos, "Learned Perceptual Image Patch Similarity (LPIPS)," *CVPR*, 2018.
32. H. Zhang *et al.*, "Residual Dense Network for Image Restoration," *IEEE Trans. Pattern Anal. Mach. Intell.*, 2019.
33. X. Chen, L. Liang, X. Guo, and K. Huang, "Deep Plug-and-Play Prior for Image Restoration," *NeurIPS*, 2019.
34. X. Liang *et al.*, "SwinIR: Image Restoration Using Swin Transformer (SwinIR)," *ICCV Workshops*, 2021. CVF Open Access
35. C. Saharia *et al.*, "Image Super-Resolution via Iterative Refinement (SR3)," *arXiv*, 2021. arXiv
36. P. Wang, L. Zhang, Z. Wang, and M. Wang, "Perceptual GAN for Single Image Super-Resolution," *ICCV*, 2017.

37. L. Yuan, X. Liu, and Y. Zhou, "A Survey of Image Super-Resolution: From Traditional Methods to Deep Learning," *ACM Computing Surveys*, 2020.
38. X. Wang, L. Zhang, Y. Liang, and K. He, "Real-ESRGAN: Training Real-World Blind Super-Resolution With Pure Synthetic Data," *ICCV Workshops*, 2021. Semantic Scholar
39. K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a Gaussian Denoiser: Residual Learning of Deep CNN for Image Denoising (DnCNN)," *TIP*, 2017.
40. L. Yuan, X. Liu, J. Wang, and Y. Xu, "Blind Image Super-Resolution with Unknown Blur Kernel," *CVPR*, 2019.
41. J. Timofte, E. Agustsson, L. Van Gool, M. Pollefeys, and L. Van Gool, "NTIRE 2019 Challenges on Image Super-Resolution: Methods and Results," *CVPR Workshops*, 2019.
42. D. Zou, H. Zhang, and J. Yuan, "KernelGAN and ZSSR for Blind Super-Resolution," *ICCV*, 2019.
43. A. Shocher, N. Cohen, and M. Irani, "Zero-Shot Super-Resolution Using Deep Internal Learning (ZSSR)," *CVPR*, 2018.
44. J. Gu, Z. Wang, and X. Wang, "Blind Super-Resolution via Spatially Variant Degradations," *ICCV*, 2019.
45. R. Timofte, V. De Smet, and L. Van Gool, "A+: Adjusted Anchored Neighborhood Regression for Fast Super-Resolution," *ACC*, 2014.
46. R. Timofte *et al.*, "NTIRE 2020 Challenge on Real Image Super-Resolution: Methods and Results," *CVPR Workshops*, 2020.
47. Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Structural Similarity (SSIM) Index for Image Quality Assessment," *TIP*, 2004.
48. R. Zeyde, M. Elad, and M. Protter, "On Single Image Scale-Up Using Sparse Representations," *Curves and Surfaces*, 2012.
49. D. Ulyanov, A. Vedaldi, and V. Lempitsky, "Deep Image Prior," *CVPR*, 2018.
50. E. Meinhardt, M. Hosseini, and M. Fritz, "Learning to Combine Complementary Restoration Tasks for Image Super-Resolution," *CVPR*, 2020.
51. S. Gu, A. Centeno, and H. Chang, "Comprehensive Benchmarking of Super-Resolution Methods," *IEEE Access*, 2021.
52. L. Wang, S. Zuo, and L. Zhang, "Adaptive Gradient Regularization for Single Image Super-Resolution," *TIP*, 2018.
53. H. Zhang, V. Sindagi, and V. M. Patel, "Image Degradation Restoration Using Deep Generative Models (Survey)," *IEEE Signal Processing Magazine*, 2022.
54. M. Saharia *et al.*, "Palette: Image-to-Image Diffusion Models for Colorization and Beyond," *NeurIPS*, 2022.
55. N. Yu *et al.*, "High-Resolution Image Synthesis with Latent Diffusion Models (LDM)," *CVPR*, 2022.
56. S. Lan, J. Markham, and N. Trigoni, "Faster Super-Resolution Networks for Edge Devices," *ECCV Workshop*, 2020.
57. C. Haris, G. Shakhnarovich, and N. Ukita, "Deep Back-Projection Networks for Super-Resolution (DBPN)," *CVPR*, 2018.
58. S. Wang, H. Fan, and J. Shi, "HAT: High-Performance Attention Transformer for Image Restoration," *ICCV*, 2023.
59. R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, "High-Resolution Image Synthesis with Latent Diffusion Models," *CVPR*, 2022.

60. K. He *et al.*, "Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification," *ICCV*, 2015.