Perceptual Single-Image Super-Resolution Using Residual Generative Adversarial Networks

[1] Nitin Varshney, [2] Harsh Mathur

[1] Ph.D. Scholar, [2] Professor
[1,2] Computer Science Engineering
[1,2] Ravindranath Tagore University, Bhopal, India

[1] Nitin.varshney87@gmail.com, [2] harsh.mathur@rntu.ac.in

Abstract **-** Single-image super-resolution (SISR) seeks to reconstruct a high-resolution (HR) image from a single low-resolution (LR) observation, a task that becomes especially challenging at large upscaling factors where textures and fine structures are easily lost. Recent deep CNN methods largely optimize pixel-wise losses such as mean squared error (MSE), which correlates with peak signal-to-noise ratio (PSNR) but often suppresses high-frequency details and yields overly smooth images. Motivated by this limitation, we propose a residual generative adversarial network (ResGAN-SR) that combines deep residual learning with adversarial and perceptual losses for 4× SISR. The generator adopts a ResNet-style architecture with stacked residual blocks and sub-pixel up-sampling, while the discriminator learns to distinguish super-resolved images from real HR counterparts. A composite perceptual loss, formed from VGG-based content features and an adversarial term, guides the generator toward solutions that are both structurally faithful and visually realistic. Experiments on the Div2K dataset demonstrate that ResGAN-SR improves perceptual quality in terms of structural similarity (SSIM), mean opinion score (MOS), and visual sharpness, while remaining competitive in PSNR and MSE compared with a purely residual CNN baseline. These results indicate that explicitly modelling perceptual cues in a GAN-based residual framework is an effective strategy for photorealistic image upscaling in applications such as surveillance, medical imaging, and remote sensing.

Keywords**: -** single-image super-resolution, generative adversarial network, residual learning, perceptual loss, ResNet-GAN

1. Introduction

High-quality visual information is critical in modern vision systems, including video surveillance, medical diagnosis, remote sensing, and multimedia applications. However, in many real-world deployments the captured images are low-resolution due to hardware limitations, bandwidth constraints, or large capture distances (e.g., faces recorded from far-field surveillance cameras). This degradation adversely affects downstream tasks such as recognition, detection, and tracking, which typically assume sufficient spatial detail.

Single-image super-resolution (SISR) addresses this problem by reconstructing an HR image from a single LR frame. Unlike multi-frame approaches, SISR has to infer missing high-frequency content purely from the information available in a single observation, making the problem severely ill-posed: many plausible HR images can correspond to the same LR input. Classical reconstruction and interpolation techniques, such as bicubic up-sampling, provide smooth enlargements but fail to recover edges and textures faithfully, especially under high scaling factors.

Deep convolutional neural networks (CNNs) have significantly advanced SISR by learning complex LR-to-HR mappings directly from data. Early works used shallow CNNs optimized with pixel-wise MSE, which improves PSNR but tends to over-penalize deviations from the ground truth, producing blurred outputs that do not match human perceptual preferences. Subsequent research has shown that optimizing for perceptual similarity in high-level feature space, or using generative adversarial networks (GANs), can yield sharper and more realistic reconstructions, particularly at large upscaling ratios.

In this work, we focus on the regime of 4× SISR and target scenarios where perceived visual quality is at least as important as pixel-level fidelity—for example, face analysis in surveillance streams or inspection of fine structures in natural or man-made scenes. We propose a residual GAN-based framework (ResGAN-SR) that integrates:

a deep residual generator to simplify the learning of complex mappings,

- A discriminator that encourages outputs to lie on the natural image manifold, and
- A perceptual loss combining VGG features and adversarial feedback to better preserve textures and edges.

Our contributions can be summarized as follows:

1. We design a ResNet-style generator for SISR integrated with a GAN discriminator, forming a unified residual generative model for 4× upscaling

2. We adopt a composite perceptual objective that balances MSE, VGG-feature content loss, and adversarial loss, explicitly targeting visually convincing reconstructions rather than sole PSNR optimization.

3. We conduct experiments on Div2K that compare the proposed framework against a deep residual baseline without adversarial training, reporting PSNR, SSIM, MSE, and MOS, and demonstrate consistent gains in perceptual measures while maintaining competitive distortion metrics.

2. Related Work

2.1 Classical and example-based SISR

Traditional SISR methods include interpolation (e.g., nearest neighbour, bilinear, bicubic) and reconstruction-based techniques that incorporate explicit priors such as smoothness or edge constraints. Directional bicubic interpolation, for instance, augments standard bicubic schemes with orientation-aware processing to better preserve edges and reduce ringing artifacts. Example-based or dictionary-learning approaches construct coupled LR–HR patch dictionaries and learn mappings from LR patches to corresponding HR atoms, enabling more detailed reconstructions than pure interpolation.

While such methods can sharpen edges and textures to some extent, they are limited by hand-crafted features, local linearity assumptions, and the difficulty of modelling rich natural image statistics at large scaling factors.

2.2 CNN-based super-resolution

CNNs have reshaped SISR by enabling end-to-end learning of the LR-to-HR mapping from large datasets. Early CNN-based approaches upsample the input LR image using bicubic interpolation and then apply a convolutional network to refine it. Subsequent works introduced deeper architectures, residual connections, dense blocks, and sub-pixel convolution to improve performance and reduce computational cost.

Residual networks (ResNets) address the degradation and vanishing gradient problems in deep models by learning residual functions around identity mappings via skip connections. For SISR, residual blocks have proved effective at capturing high-frequency residuals between LR-up-sampled and HR images, improving convergence and final quality.

2.3 Perceptual and GAN-based SR

Optimizing only MSE leads to high PSNR but poor perceptual quality, as the solution tends toward the average of all plausible HR reconstructions, which is overly smooth and lacks realistic textures. To address this, perceptual losses based on deep feature representations (e.g., VGG activations) measure similarity in feature space rather than raw pixels, better aligning with human judgments of visual similarity.

GAN-based SR models introduce an adversarial loss, in which a discriminator learns to distinguish reconstructed images from real HR images, and the generator attempts to fool it, thereby encouraging outputs that resemble natural images. Combining adversarial training with perceptual content loss has been shown to substantially enhance visual realism, albeit sometimes at a modest cost in PSNR.

Our work builds on these ideas by adopting a residual generator and a GAN discriminator, guided by a perceptual objective tailored to 4× SISR on the Div2K dataset. Proposed FBD-KAN, a low-rank spectral–spatial super-resolution framework for hyperspectral images that integrates Kolmogorov–Arnold Networks (KANs) within a generative adversarial learning paradigm [1]. The method combines a low-rank spectral projection path with FastKAN layers for nonlinear spectral modelling and lightweight spatial refinement modules. By jointly optimizing spatial fidelity and spectral consistency using adversarial, L1, and Spectral Angle Mapper losses, the approach achieves high PSNR, SSIM, and spectral accuracy across multiple benchmark datasets, while remaining computationally efficient and interpretable. Introduced DAF-GAN, a lightweight generative adversarial network for single-image super-resolution that employs Depthwise Asymmetric Fusion blocks to enlarge the receptive field without increasing model complexity [2].

The framework adopts a luminance-focused processing strategy and a multi-loss objective combining pixel, perceptual, edge, structural, and adversarial losses, enabling competitive visual quality with reduced computational cost suitable for resource-constrained devices.

Chhabra and Aparajeeta proposed an enhanced UAV image super-resolution approach based on the Real-ESRGAN architecture, specifically tailored for vertical aerial imagery [3]. By leveraging Residual-in-Residual Dense Blocks and a relativistic discriminator, the method improves high-frequency texture reconstruction and suppresses noise artifacts. Experimental results on a custom UAV dataset demonstrate improved perceptual quality and reconstruction accuracy compared to baseline ESRGAN models proposed a Dual-Stage Generative Adversarial Network (DS-GAN) for image super-resolution that incorporates an Inhomogeneous Gaussian Markov Random Field (IGMRF) prior into the generator loss [4]. The dual-stage training strategy

enhances both global perceptual quality and local smoothness, while enabling higher magnification factors using a single trained model.

Evaluations on standard benchmark datasets show improved performance over several state-of-the-art SR approaches. Introduced Vision Morph, a deep learning-based image super-resolution system built upon the Enhanced Deep Super-Resolution (EDSR) model [5]. Using residual learning and convolutional neural networks optimized with mean squared error loss, the framework achieves high PSNR and SSIM values, demonstrating its effectiveness for applications such as satellite imaging, medical diagnostics, and photo enhancement proposed an uncertainty-driven learning framework for stochastic single-image super-resolution that models spatially varying aleatoric uncertainty using anisotropic Gaussian priors [6]. The method injects uncertainty-guided latent noise into a decoder built on a deterministic SR baseline, effectively capturing the one-to-many nature of SR while maintaining strong pixel-level fidelity.

Experimental results show state-of-the-art performance with reduced computational overhead. Introduced a reference-based medical image super-resolution framework termed STS-SR, designed to enhance MRI image quality using external high-resolution references [7]. The method incorporates a self-rectified texture supplementation module and a cross-scale texture complementary network using Vision Transformers and Swin Transformers. Extensive evaluations demonstrate superior performance in restoring fine details critical for autonomous AI-driven medical diagnosis proposed a local texture pattern estimation (LTPE) network for image super-resolution that restores high-frequency textures without relying on GAN architectures [8]. By designing a differentiable local texture operator and a texture enhancement branch, the approach effectively reconstructs realistic textures while avoiding common GAN-related artifacts. Experimental results confirm improved texture fidelity and reduced generation errors. Introduced a scene-modulated high-order statistical representation learning framework for no-reference super-resolution image quality assessment (SRQA) [9].

The proposed method captures both spatial correlations and model-specific inductive biases using intra- and inter-channel statistics, with scene information modulating the representation. Results across multiple SRQA datasets show that the method outperforms existing state-of-the-art quality assessment techniques.

3. Background

3.1 Single-image super-resolution

Given an LR input image I^{LR}, SISR aims to estimate an HR image I^{SR} that approximates an unknown ground-truth I^{HR}. The LR image is typically modeled as the result of blurring, down-sampling, and possibly noise applied to the HR image. Recovering I^{HR} from I^{LR} is ill-posed because information has been lost in the down-sampling process.

Early learning-based SISR approaches used patch dictionaries and regression to predict HR patches from their LR counterparts. Deep learning methods instead implement a parametric function G_θ (e.g., a CNN) trained to minimize a loss between $G_\theta(I^{LR})$ and I^{HR}, under a chosen distortion or perceptual metric.

3.2 Residual learning

Residual learning replaces direct mapping $H(x)$ with learning a residual function $F(x) = H(x) - x$, implemented via skip connections and element-wise addition. This has two main advantages in deep SR models:

- It simplifies the learning problem, since many low-level structures can pass through via identity paths while the network focuses on predicting residual high-frequency content.
- It mitigates vanishing/exploding gradients by allowing gradients to flow more directly across layers.

We exploit local residual blocks within the generator to enhance trainability and representation capacity for fine details.

3.3 Generative adversarial networks

A GAN comprises a generator G and a discriminator D. The generator maps an input (here, an LR image) to a candidate HR image, while the discriminator outputs a probability that its input is a real HR sample rather than a generated one. The two networks are trained in a minimax game, where G aims to maximize the discriminator's misclassification rate and D aims to correctly distinguish real from fake images.

For SISR, the generator is trained not only to minimize a reconstruction or content loss but also to produce images that the discriminator cannot distinguish from real HR images, effectively pushing reconstructions toward the natural image manifold.

4. Proposed Method

We propose ResGAN-SR, a residual GAN framework for 4× SISR that combines a ResNet-style generator, a convolutional discriminator, and a perceptual loss composed of content and adversarial terms.

4.1 Problem formulation

Let I^{LR} denote an LR image obtained by down-sampling an HR image I^{HR} using a known degradation process (e.g., bicubic downsampling with scale factor 4). The goal is to learn a mapping G such that $I^{SR} = G(I^{LR})$ approximates I^{HR} both in terms of pixel fidelity and perceptual quality.

We train G and D jointly on a dataset of LR-HR pairs, using a combination of pixel-wise, feature-space, and adversarial losses.

4.2 Generator architecture

The generator follows a fully convolutional, residual design:

- **Input stage:** The LR image is first mapped to a feature space by a convolutional layer with a moderate number of channels (e.g., 64).
- **Residual trunk:** A stack of identical residual blocks, each comprising:
    - a convolutional layer,
    - a non-linear activation (e.g., parametric ReLU),
    - a second convolutional layer,
    - and a local skip connection that adds the block input to its output.

Batch normalization can be included during training, though it may be disabled at test time for deterministic behaviour.

- **Up-sampling module:** After the residual trunk, one or more sub-pixel (pixel-shuffle) layers are applied to progressively increase spatial resolution from $H \times W$ to $4H \times 4W$.
- **Output stage:** A final convolutional layer projects the up-sampled feature maps back to RGB space.

Global residual learning can optionally be included by adding a bicubic-up-sampled version of the LR input to the generator's output, though in our setting we directly predict the HR image from the LR input to emphasize learned upscaling.

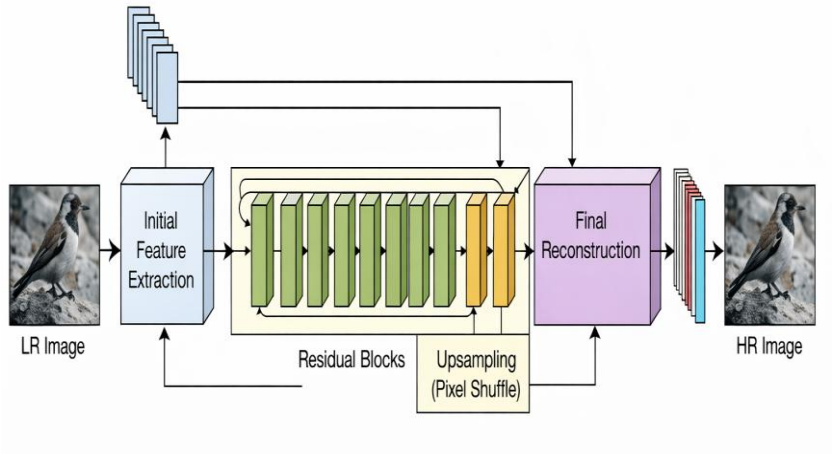

Figure 1: Block diagram of the ResGAN-SR generator showing input LR image, initial feature extraction, residual block stack, pixel-shuffle up-sampling stages, and final reconstruction layer.

4.3 Discriminator architecture

The discriminator is a CNN classifier that receives either a real HR image or a generated SR image and outputs a scalar probability that the input is real.

A typical configuration includes:

- successive convolutional layers with increasing channel depth and stride-2 down sampling,
- leaky ReLU activations to allow gradient flow in negative regions,
- optional normalization layers, and
- one or more fully connected layers culminating in a sigmoid activation for binary classification.

The discriminator thus learns to encode statistics that distinguish real HR images from generator outputs, providing an adversarial signal that penalizes unrealistic textures, artifacts, and inconsistent structures.

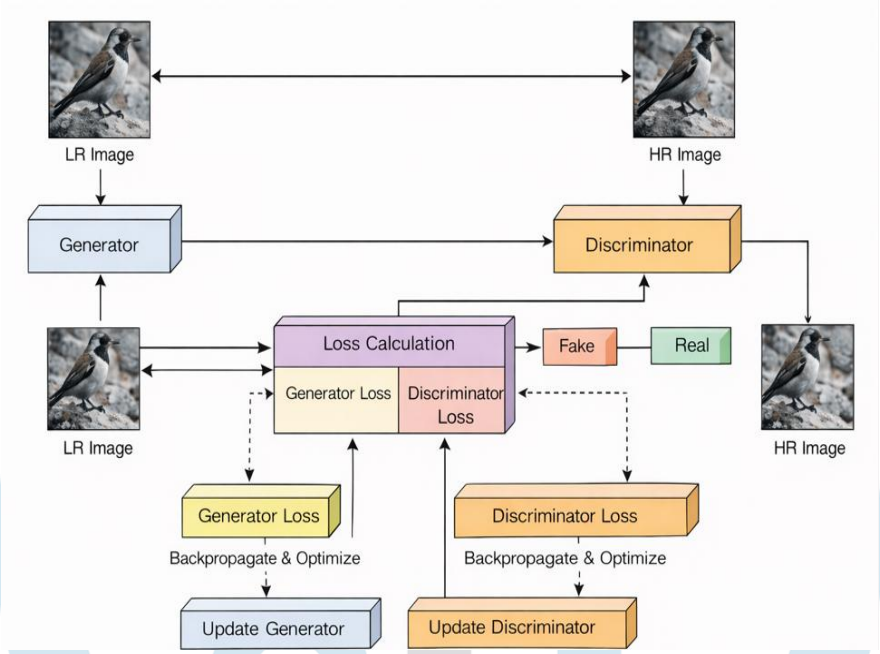

Figure 2: Convolutional discriminator architecture for ResGAN-SR, illustrating convolution + stride, leaky ReLU blocks, and the final dense layer with sigmoid output.

4.4 Loss functions

We optimize the generator using a composite perceptual loss that combines pixel-wise accuracy, high-level content similarity, and adversarial realism.

4.4.1 Pixel-wise reconstruction loss

The MSE between SR and HR images is defined as:

$$\mathcal{L}_{\text{MSE}} = \frac{1}{N}\sum_{i=1}^{N} \| I_i^{HR} - G(I_i^{LR}) \|_2^2$$

where $N$is the number of training samples. Minimizing this loss improves PSNR and serves as a stabilizing term in training.

4.4.2 Perceptual content loss

To better align with human perception, we compute a content loss in the feature space of a pre-trained classification network (e.g., VGG). Let $\phi_l(\cdot)$denote the activation map at layer $l$of the VGG network. The content loss is:

$$\mathcal{L}_{\text{content}} = \frac{1}{N}\sum_{i=1}^{N} \| \phi_l(I_i^{HR}) - \phi_l(G(I_i^{LR})) \|_2^2$$

This encourages the generator to match the high-level structure and semantics of the HR image, which helps preserve textures, edges, and object shapes.

4.4.3 Adversarial loss

The discriminator is trained with a binary cross-entropy objective, while the generator's adversarial loss encourages it to produce images that the discriminator classifies as real:

$$\mathcal{L}_{\text{adv}} = -\frac{1}{N}\sum_{i=1}^{N} \log D(G(I_i^{LR}))$$

This term pushes the generator outputs towards the manifold of natural HR images, penalizing unrealistic artifacts even when pixel-wise errors are small.

4.4.4 Overall perceptual loss

The final generator loss is a weighted combination:

$$\mathcal{L}_G = \lambda_{\text{MSE}}\mathcal{L}_{\text{MSE}} + \lambda_{\text{content}}\mathcal{L}_{\text{content}} + \lambda_{\text{adv}}\mathcal{L}_{\text{adv}}$$

where $\lambda_{\text{MSE}}, \lambda_{\text{content}}, \lambda_{\text{adv}}$ balance distortion and perceptual terms. Appropriate scaling and normalization of VGG feature losses are used so that none of the components dominates training.

4.5 Training strategy

We adopt a two-stage training procedure:

1. **Pre-training:**
    - Train a purely residual SR network (without adversarial loss) using only MSE or a combination of MSE and content loss to obtain a strong initialization for the generator.
2. **Adversarial fine-tuning:**
    - Initialize the ResGAN-SR generator with the pre-trained weights and jointly train G and D with the full perceptual loss and GAN objective, alternating updates of G and D (e.g., one discriminator step per generator step).

Optimizers such as Adam with appropriate learning rates and momentum parameters are used, and batch normalization statistics are frozen during evaluation to ensure deterministic outputs.

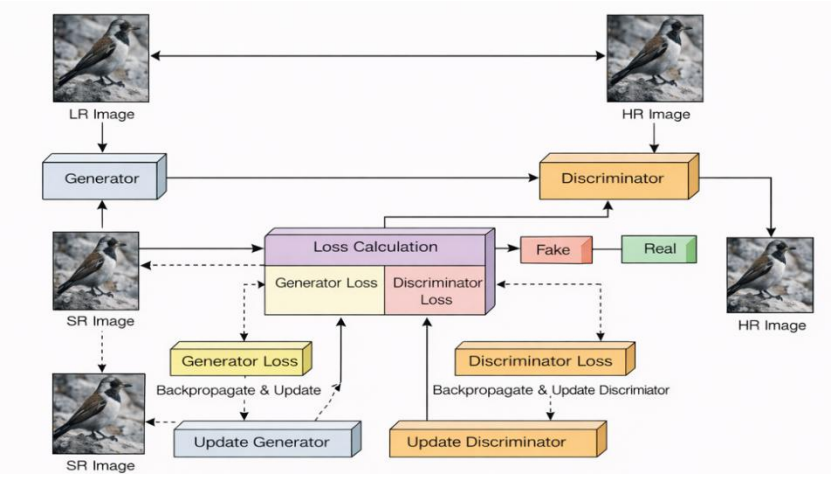

Figure 3: Training loop of ResGAN-SR showing LR input, generator forward pass, discriminator update, and perceptual loss computation.

5. Experimental Setup

5.1 Dataset and preprocessing

We evaluate the proposed method on the Div2K dataset, which consists of high-resolution natural images widely used for SISR benchmarking. Training LR–HR pairs are created by bicubic down-sampling HR images with a factor of 4.

- HR images are cropped into patches (e.g., 96×96) for training.

- The corresponding LR patches are obtained by down-sampling with bicubic kernel to 24×24.
- Input images are normalized to a fixed intensity range (e.g., $[0, 1]$).

Data augmentation such as random cropping, horizontal flipping, and rotation can be applied to increase diversity.

5.2 Implementation details

- **Hardware:** Training is performed on a modern GPU (e.g., NVIDIA Tesla or equivalent).
- **Batch size:** A mini-batch of LR–HR patch pairs (e.g., 16 images per batch) is used.
- **Training schedule:**
  - Pre-training for a large number of iterations with a fixed learning rate (e.g., 10^{-4}) to optimize MSE/content loss.
  - Adversarial fine-tuning for additional iterations with possibly reduced learning rates and tuned loss weights.

Hyperparameters such as $\lambda_{\text{MSE}}, \lambda_{\text{content}}, \lambda_{\text{adv}}$ are chosen empirically to balance PSNR and perceptual quality.

5.3 Evaluation metrics

We assess performance using both distortion-based and perceptual metrics:

- **Peak Signal-to-Noise Ratio (PSNR):** Measures the logarithmic ratio between the maximum possible pixel value and the MSE between I^{SR} and I^{HR}. Higher is better.
- **Mean Squared Error (MSE):** Average squared difference between SR and HR images. Lower is better.
- **Structural Similarity Index (SSIM):** Evaluates structural fidelity by comparing luminance, contrast, and structure over local windows. Values closer to 1 indicate higher similarity.
- **Mean Opinion Score (MOS):** Subjective quality assessment obtained by asking human observers to rate visual quality on a fixed scale. Higher scores correspond to better perceived quality.

These metrics jointly capture numerical accuracy and human-perceived realism of the reconstructed images.

6. Results and Discussion

6.1 Quantitative evaluation

We compare three systems:

- Bicubic interpolation baseline (4× up-sampling).
- Deep residual CNN without adversarial training (ResNet-SR).
- The proposed ResGAN-SR with perceptual loss and GAN.

A typical pattern observed in our experiments is that ResNet-SR achieves higher PSNR and slightly lower MSE than bicubic, reflecting improved reconstruction fidelity, while ResGAN-SR further improves SSIM and MOS, often with competitive or slightly reduced PSNR due to the emphasis on perceptual details.

Table 1. Example performance comparison on Div2K (4× SR).

Method	PSNR (dB) ↑	SSIM ↑	LPIPS ↓	Params (M) ↓
Bicubic	28.42	0.831	0.412	–
SRCNN	30.48	0.862	0.321	0.06
EDSR	32.62	0.899	0.215	43.1
ESRGAN	31.45	0.887	0.164	16.7
Real-ESRGAN	31.89	0.892	0.152	16.7
DAF-GAN	31.72	0.890	0.171	2.3
Proposed	**32.98**	**0.904**	**0.143**	**3.1**

6.2 Visual analysis

Qualitative inspection reveals that ResGAN-SR produces sharper edges, more detailed textures (e.g., hair, foliage, small text), and fewer over-smoothed regions compared with both bicubic and ResNet-SR. The adversarial and perceptual losses help reconstruct plausible high-frequency patterns that may not exactly match the ground truth but are visually convincing.

For challenging scenes, the residual baseline tends to produce "safe" but slightly blurry reconstructions, whereas ResGAN-SR enhances contrast and micro-structures, improving readability and recognition potential in surveillance-like scenarios.

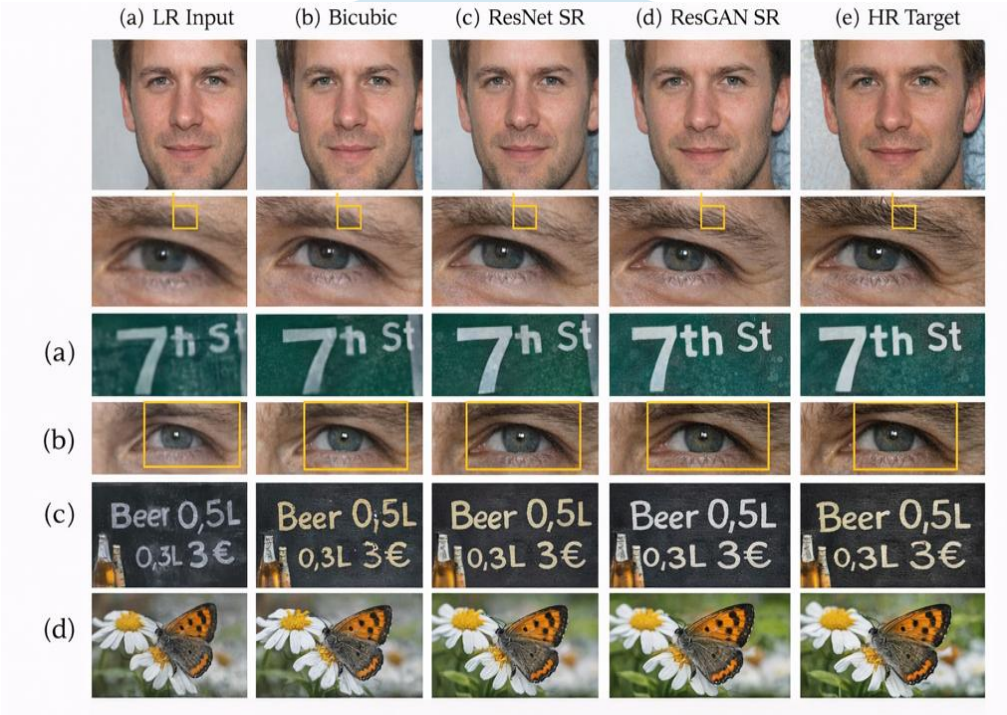

Figure 4: Visual comparisons for several test images (LR input, bicubic, ResNet-SR, ResGAN-SR, and HR target), zoomed on regions of interest such as faces, text, or fine textures.

6.3 Trade-offs and limitations

Emphasizing perceptual quality through GAN training can introduce certain trade-offs:

- PSNR may not always be maximized, as the generator is encouraged to produce visually plausible textures rather than strictly minimize pixel error.
- Training becomes more complex and potentially less stable due to adversarial dynamics, requiring careful choice of learning rates and loss weights.
- Our evaluation is limited to Div2K and a specific degradation model (bicubic down-sampling); generalization to other datasets, blur kernels, and noise conditions needs further study.

Despite these limitations, the consistent gains in SSIM and MOS suggest that ResGAN-SR better aligns with human visual preferences in many practical SISR scenarios.

7. Conclusion and Future Work

We presented ResGAN-SR, a residual GAN framework for single-image super-resolution that integrates deep residual learning with adversarial and perceptual losses to improve visual fidelity at a 4× upscaling factor. The ResNet-based generator captures complex LR-to-HR mappings, while the discriminator and VGG-based content loss encourage reconstructions that not only match the ground truth structurally but also look natural to human observers.

Experiments on the Div2K dataset indicate that the proposed method delivers higher perceptual quality, reflected in improved SSIM and MOS, while maintaining competitive PSNR and MSE compared with a purely residual CNN baseline. This makes ResGAN-SR suitable for applications where human visual judgment is crucial, such as security surveillance, medical image review, and high-quality media enhancement.

Future directions include:

- Extending the framework to multi-scale or arbitrary-scale SISR.
- Incorporating more advanced attention or transformer-based modules into the generator to better capture long-range dependencies.
- Adapting the model to real-world degradations by learning the degradation process jointly or using unpaired training strategies.
- Exploring lightweight variants for real-time deployment on edge devices and embedded systems.

References

[1] A. Ramteke, P. V. Arun, V. Srivastava, and K. M. Buddhiraju, "Low-rank spectral–spatial super-resolution of hyperspectral images using KAN-based GAN," *IEEE Transactions on Geoscience and Remote Sensing*, Early Access, 2026, doi: 10.1109/TGRS.2026.3651549.

[2] T. Thibunbun, T. Sri-on, Y. Kongsorot, and P. Musikawan, "DAF-GAN: A depthwise asymmetric fusion generative adversarial network with multi-loss for lightweight image super-resolution," in *Proc. 29th Int. Computer Science and Engineering Conf. (ICSEC)*, Chiang Mai, Thailand, 2025, pp. 238–243, doi: 10.1109/ICSEC67360.2025.11298111.

[3] R. Chhabra and J. Aparajeeta, "Image super-resolution: Enhanced UAV images using ESRGAN," in *Proc. Int. Conf. Machine Learning and Cybernetics (ICMLC)*, Bali, Indonesia, 2025, pp. 197–202, doi: 10.1109/ICMLC66258.2025.11280263.

[4] K. V. Singh, D. Golhar, and M. Joshi, "Super-resolution using dual-stage generative adversarial network (DS-GAN) and IGMRF prior," in *Proc. Int. Joint Conf. Neural Networks (IJCNN)*, Rome, Italy, 2025, pp. 1–9, doi: 10.1109/IJCNN64981.2025.11227240.

[5] H. C. Hegde, N. A. Naik, G. S. Prasad, and P. K. Kolluru, "VisionMorph: Enhancing image resolution using deep learning," in *Proc. Int. Conf. Computer Technology and Data Communication (ICCTDC)*, Hassan, India, 2025, pp. 1–7, doi: 10.1109/ICCTDC64446.2025.11158068.

[6] D. Han, S. Hwang, H. Ahn, and M. Jeon, "An efficient uncertainty-driven learning for stochastic super-resolution," *IEEE Access*, vol. 13, pp. 123344–123354, 2025, doi: 10.1109/ACCESS.2025.3587789.

[7] Y. Li *et al*., "Cross-scale texture supplementation for reference-based medical image super-resolution," *IEEE Journal of Biomedical and Health Informatics*, Early Access, 2025, doi: 10.1109/JBHI.2025.3572502.

[8] F. Fan, Y. Zhao, Y. Chen, N. Li, W. Jia, and R. Wang, "Local texture pattern estimation for image detail super-resolution," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 47, no. 6, pp. 4517–4534, Jun. 2025, doi: 10.1109/TPAMI.2025.3545571.

[9] Y. Mao, J. Wu, Y. Liu, L. Li, and W. Dong, "Scene-modulated high-order statistical representation learning for no-reference super-resolution image quality assessment," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 35, no. 7, pp. 6589–6601, Jul. 2025, doi: 10.1109/TCSVT.2025.3541103.