A Comprehensive Analysis of Domain Adaptation Techniques in Image Super-Resolution

Meripe Sathvik
Roll No: 220010031
Email: 220010031@iitdh.ac.in

Janapati Mahesh
Roll No: 220010022
Email: 220010022@iitdh.ac.in

Akshay Pasupuleti
Roll No: 220010041
Email: 220010041@iitdh.ac.in

***Abstract*—Image super-resolution (SR) aims to reconstruct high-resolution (HR) images from low-resolution (LR) inputs. This paper provides a comprehensive analysis of domain adaptation techniques in SR, focusing on three state-of-the-art methods: DASRGAN for infrared image super-resolution, an unsupervised real-world SR approach, and the Source-Free Adaptive Image Super-Resolution with Wavelet Augmentation Transformer (SODA-SR). These methods leverage domain adaptation strategies to enhance SR performance in various domains, addressing challenges such as noise suppression, texture enhancement, and adaptation to unseen degradations without relying on source data. Key contributions include texture-oriented and noise-oriented adaptation, feature alignment through adversarial training, wavelet-based augmentation, uncertainty-aware self-training, and frequency-domain regularization losses. Experimental results demonstrate the effectiveness of these techniques in achieving superior SR performance across multiple datasets, highlighting the significant impact of domain adaptation in advancing SR capabilities.**

***Index Terms*—Super-Resolution, Domain Adaptation, Infrared Imaging, Unsupervised Learning, Feature Alignment, Wavelet Augmentation Transformer, Uncertainty-Aware Self-Training, Computer Vision.**

I. Introduction

Image super-resolution (SR) is a fundamental task in computer vision, aiming to reconstruct high-resolution (HR) images from low-resolution (LR) counterparts. It plays a critical role in various applications, including medical imaging, satellite imagery, surveillance, and consumer electronics, where obtaining high-quality images is essential but often constrained by hardware limitations, cost, or environmental factors.

Traditional SR methods have achieved remarkable success using deep learning techniques, particularly convolutional neural networks (CNNs) and generative adversarial networks (GANs). However, these methods typically rely on supervised learning with paired LR-HR datasets, often synthesized using simple degradation models like bicubic downsampling. Such synthetic datasets fail to capture the complex and diverse degradations present in real-world scenarios, leading to poor generalization and suboptimal performance when applied outside of the training domain.

Domain adaptation techniques have emerged as a promising solution to bridge the gap between synthetic training data and real-world applications. By transferring knowledge from a source domain to a target domain, domain adaptation reduces the performance degradation caused by domain shifts and enables models to generalize better to unseen data.

This paper explores three innovative SR methods that leverage domain adaptation to address the challenges of real-world SR tasks:

1) **DASRGAN for Infrared Image Super-Resolution** [1]: A framework that enhances texture fidelity and suppresses noise in infrared (IR) images through domain adaptation strategies involving texture-oriented and noise-oriented modules.
2) **Unsupervised Real-World Super-Resolution** [2]: An approach that addresses the domain gap between synthetic and real-world LR images by aligning feature distributions via adversarial training, enabling SR without the need for paired datasets.
3) **Source-Free Adaptive Image Super-Resolution with Wavelet Augmentation Transformer (SODA-SR)** [3]: A novel Source-Free Domain Adaptation (SFDA) framework that leverages pseudo-labels and wavelet-based augmentation to bridge the domain gap without accessing source data.

By thoroughly analyzing these methods, we aim to provide insights into how domain adaptation can be effectively employed in SR tasks across different domains, ultimately contributing to the advancement of image restoration techniques. The remainder of this paper is organized as follows: Section II discusses the background and related work, Section III presents the proposed methods, Section IV details the experimental results, Section V provides a comparative analysis, Section VI outlines future work, and Section VII concludes the paper.

II. Background and Related Work

A. Super-Resolution Techniques

Image super-resolution has been extensively studied, with early methods relying on interpolation techniques such as nearest-neighbor, bilinear, and bicubic interpolation. These methods are simple but often produce blurry results lacking fine details.

The advent of deep learning brought significant advancements in SR. Dong et al. introduced SRCNN [4], a pioneering work that applied CNNs to SR, achieving superior results compared to traditional methods. Subsequent models, such as

VDSR [5], EDSR [6], and RCAN [7], pushed the state-of-the-art by employing deeper networks, residual connections, and attention mechanisms.

Generative Adversarial Networks (GANs) [8] have further revolutionized SR by enabling the generation of perceptually realistic images. SRGAN [9] combined adversarial training with perceptual losses to produce high-fidelity images. ESRGAN [10] improved upon this by introducing residual-in-residual dense blocks and a relativistic discriminator, leading to more natural textures and sharper details.

Despite these successes, most SR models are trained on synthetic datasets with simplistic degradation models, limiting their applicability to real-world scenarios with complex and unknown degradations.

B. Domain Adaptation in Super-Resolution

Domain adaptation seeks to transfer knowledge from a source domain (e.g., synthetic data) to a target domain (e.g., real-world data) by reducing the domain shift. In SR, domain adaptation addresses the gap between training on synthetic LR-HR pairs and deploying models on real-world LR images.

Unsupervised domain adaptation methods, such as CycleGAN [11], enable learning from unpaired data by leveraging cycle consistency and adversarial losses. In SR, approaches like BSRGAN [12] simulate more realistic degradations to improve generalization. However, these methods may not fully capture the diversity of real-world degradations or handle specific domain challenges like noise in infrared images.

C. Infrared Image Super-Resolution

Infrared imaging captures thermal radiation, providing critical information in applications such as surveillance, night vision, and medical diagnostics. IR images often suffer from low resolution and high noise levels due to hardware limitations and environmental factors. Traditional SR models struggle with IR images because they do not account for the unique noise characteristics and lack of visible light information.

Recent works have attempted to address IR SR by incorporating domain-specific knowledge or specialized architectures [13]. However, challenges remain in effectively enhancing textures while suppressing noise, particularly in the absence of large-scale paired datasets.

III. Proposed Methods

A. DASRGAN for Infrared Image Super-Resolution

1) Overview: The Domain Adaptation Super-Resolution Generative Adversarial Network (DASRGAN) [1] addresses the challenges of infrared image super-resolution (SR), focusing on noise suppression and texture enhancement. Unlike visible light images, IR images often exhibit high levels of noise and lack detailed textures due to sensor limitations.

DASRGAN introduces two key adaptation modules:

- **Texture-Oriented Adaptation (TOA)**: Aligns texture distributions between source (visible light) and target (infrared) domains to enhance details.
- **Noise-Oriented Adaptation (NOA)**: Suppresses noise and minimizes artifacts during the generation process.

2) Architecture: The DASRGAN architecture consists of a generator and two discriminators, designed to balance texture enhancement and noise suppression:

- **Generator** (G): A CNN-based model that maps low-resolution (LR) infrared images to high-resolution (HR) outputs. It employs Residual-in-Residual Dense Blocks (RRDB) with skip connections for efficient gradient flow and faster convergence.
- **Texture Discriminator** (D_t): Incorporates a Sobel-based prior branch to enhance edge and texture details. It penalizes discrepancies in texture patterns, guiding the generator to produce sharper images.
- **Noise Discriminator** (D_n): Differentiates between noise patterns and desired features, helping the generator produce clean, high-quality images.

Figure 1 illustrates the complete DASRGAN architecture.

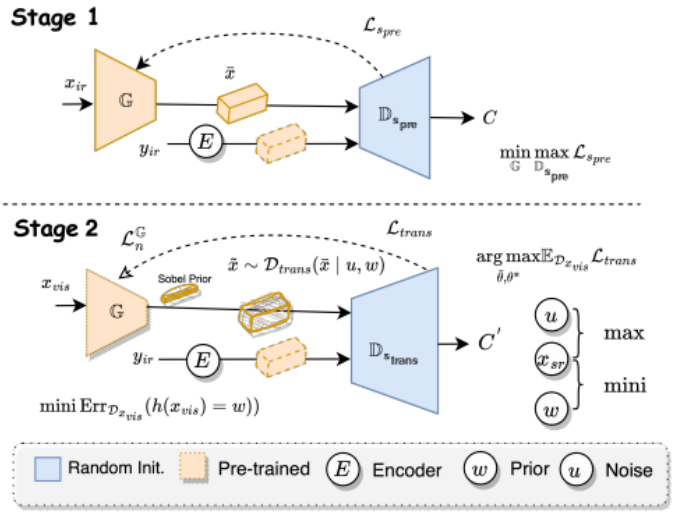


Fig. 1: DASRGAN architecture showing the generator, texture discriminator (D_t), and noise discriminator (D_n).

3) Training Strategy: DASRGAN employs a two-stage training approach:

- **Stage 1: Pre-training**
  - The generator (G) maps low-resolution (LR) IR images to high-resolution (HR) images using Mean Absolute Error (MAE) loss.
  - The discriminator (D_{spre}) distinguishes real and generated HR images using cross-entropy loss.
- **Stage 2: Target-Oriented Domain Adaptation**
  - The generator adapts visible-light LR images to generate HR IR images.
  - The texture discriminator (D_t) incorporates a Sobel-based prior branch to align texture distributions.
  - Two additional loss functions guide this stage:
    - **Texture-Oriented Prior Adversarial Loss** ($\mathcal{L}^t_{\text{adv}}$): Aligns texture distributions between domains.

∗ **Noise Adversarial Loss ($\mathcal{L}^n_{\text{adv}}$)**: Minimizes noise artifacts.

4) Loss Functions: The generator loss is defined as:

$$\mathcal{L}_G = \mathcal{L}^t_{\text{adv}} + \mathcal{L}^n_{\text{adv}} + \lambda_{\text{perc}}\mathcal{L}_{\text{perc}} + \lambda_{\text{cont}}\mathcal{L}_{\text{cont}} \tag{1}$$

where:

- $\mathcal{L}^t_{\text{adv}}$ and $\mathcal{L}^n_{\text{adv}}$: Adversarial losses from D_t and D_n:

$$\mathcal{L}^t_{\text{adv}} = \mathbb{E}_x[\log D_t(G(x))] \tag{2}$$

$$\mathcal{L}^n_{\text{adv}} = \mathbb{E}_x[\log D_n(G(x))] \tag{3}$$

- $\mathcal{L}_{\text{perc}}$: Perceptual loss using feature maps from a pre-trained VGG network:

$$\mathcal{L}_{\text{perc}} = \sum_i \|\phi_i(G(x)) - \phi_i(y)\|_2^2 \tag{4}$$

- $\mathcal{L}_{\text{cont}}$: Content loss, defined as:

$$\mathcal{L}_{\text{cont}} = \|G(x) - y\|_2^2 \tag{5}$$

Discriminators are trained with adversarial losses:

$$\mathcal{L}_{D_t} = -\mathbb{E}_y[\log D_t(y)] - \mathbb{E}_x[\log(1 - D_t(G(x)))], \tag{6}$$

$$\mathcal{L}_{D_n} = -\mathbb{E}_y[\log D_n(y)] - \mathbb{E}_x[\log(1 - D_n(G(x)))]. \tag{7}$$

5) Implementation Details: Training uses the Adam optimizer with a learning rate of 1×10^{-5}. The generator consists of residual blocks with instance normalization and LeakyReLU activation. Sobel-based texture priors improve alignment, while PatchGAN-based discriminators ensure local realism in the generated images. Hyperparameters λ_{perc} and λ_{cont} are tuned for optimal performance.

B. Unsupervised Real-World Super-Resolution

1) Overview: The unsupervised real-world SR approach [2] tackles the challenge of SR in the absence of paired LR-HR datasets, which is common in real-world scenarios due to unpredictable and complex degradations. The method employs domain adaptation techniques to bridge the domain gap between synthetic (source) and real-world (target) images by aligning feature distributions through adversarial training.

2) Architecture: The architecture of the unsupervised real-world SR method integrates an encoder-decoder network and two discriminators to achieve domain adaptation and high-quality super-resolution:

- **Encoder** (E): Extracts feature representations from low-resolution (LR) input images, capturing both global and local structures for effective reconstruction.
- **Decoder** (G): Reconstructs high-resolution (HR) images from the encoded features. It utilizes skip connections and convolutional layers to retain spatial information and improve detail preservation.
- **Feature Discriminator** (D_f): Operates at the feature level to ensure alignment between the feature distributions of the source and target domains.
- **Image Discriminator** (D_i): Works at the image level to distinguish between real HR images and generated HR outputs, enforcing perceptual and visual consistency.

Figure 2 illustrates the overall architecture, highlighting the interactions between the encoder, decoder, and discriminators.

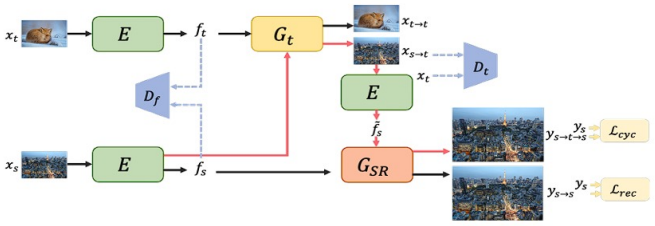


Fig. 2: Unsupervised Real-World SR architecture showing the encoder (E), decoder (G), feature discriminator (D_f), and image discriminator (D_i).

3) Methodology: The architecture consists of an encoder-decoder network and two discriminators:

- **Encoder** (E): Extracts feature representations from input LR images.
- **Decoder** (G): Reconstructs HR images from the encoded features.
- **Feature Discriminator** (D_f): Ensures that the encoded features from the target domain resemble those from the source domain.
- **Image Discriminator** (D_i): Distinguishes between generated HR images and real HR images.

The overall training process involves:

- **Feature-Level Alignment**: Using adversarial loss to align the feature distributions between source and target domains.
- **Image-Level Alignment**: Encouraging the decoder to generate images indistinguishable from real HR images.
- **Self-Supervised Learning**: Utilizing cycle consistency and identity losses to preserve content and stabilize training.

4) Loss Functions:

- **Adversarial Loss ($\mathcal{L}_{\text{adv}}$)**:

$$\begin{aligned}\mathcal{L}_{\text{adv}} = {} & \mathbb{E}_{x_t}\left[\log D_f(E(x_t))\right] \\ & + \mathbb{E}_{x_s}\left[\log\left(1 - D_f(E(x_s))\right)\right] \\ & + \mathbb{E}_{G(E(x_t))}\left[\log D_i(G(E(x_t)))\right]\end{aligned} \tag{8}$$

This loss ensures that the encoder (E) and generator (G) produce outputs indistinguishable from real data by fooling the discriminators D_f and D_i.

- **Cycle Consistency Loss ($\mathcal{L}_{\text{cyc}}$)**:

$$\mathcal{L}_{\text{cyc}} = \mathbb{E}_{x_t}\|E(G(E(x_t))) - E(x_t)\|_1 \tag{9}$$

This loss enforces that the encoded representation of a generated image maps back to the original encoding, preserving domain-specific features.

- **Identity Loss ($\mathcal{L}_{\text{id}}$)**:

$$\mathcal{L}_{\text{id}} = \mathbb{E}_{y_s}\|G(E(y_s)) - y_s\|_1 \tag{10}$$

This loss encourages the generator to retain identity mappings when the input image already belongs to the target domain, ensuring stability during training.

- **Hyperparameters (λ_{adv}, λ_{cyc}, λ_{id})**: These control the relative importance of the adversarial, cycle consistency, and identity losses.

5) Implementation Details: The encoder and decoder networks are constructed with convolutional layers, batch normalization, and ReLU activations. The discriminators use PatchGAN structures to evaluate local features. The model is trained using the Adam optimizer with carefully selected learning rates and hyperparameters.

Data augmentation techniques such as random cropping, flipping, and rotation are applied to enhance generalization. The model is trained on unpaired LR and HR images from the source and target domains, making it suitable for real-world applications where paired data is unavailable.

C. SODA-SR: Source-Free Adaptive Image Super-Resolution with Wavelet Augmentation Transformer

1) Overview: The Source-Free Domain Adaptation (SFDA) framework, named SODA-SR [3], addresses the practical scenario where access to source data is restricted due to privacy or legal constraints. SODA-SR adapts a pre-trained source model to a target domain using only unlabeled target data, without relying on source data or paired datasets.

Key contributions of SODA-SR include:

- **Wavelet Augmentation Transformer (WAT)**: Introduces wavelet-based data augmentation within the model to enhance robustness to domain shifts.
- **Uncertainty-Aware Self-Training**: Improves pseudo-label accuracy by estimating and accounting for uncertainty in predictions.
- **Frequency-Domain Regularization**: Employs regularization losses in the frequency domain to prevent overfitting and preserve important image structures.

2) Architecture: The SODA-SR framework incorporates a teacher-student model architecture to achieve source-free domain adaptation for super-resolution tasks:

- **Teacher Model**: Processes one low-resolution (LR) input image along with seven geometrically augmented versions (rotated and flipped) to generate refined pseudo-labels. The Softmax normalization in the teacher model is replaced by Gumbel-Softmax [**?**] for uncertainty-aware label generation.
- **Pseudo-Label Generation**: For each LR input image, the teacher model runs multiple times to produce N pseudo-labels. These are used to compute the mean and variance, enabling uncertainty estimation for the pseudo-labels.
- **Wavelet Augmentation Transformer (WAT)**: Captures multi-scale and multi-frequency information by leveraging discrete wavelet transforms and deformable attention mechanisms. It integrates features across different wavelet levels to enhance robustness and detail preservation in the SR outputs.
- **Student Model**: Adapts to the target domain by fine-tuning on the pseudo-labels generated by the teacher model, learning domain-specific features without relying on source data.

Figure 3 illustrates the proposed SODA-SR framework, showing the teacher-student interaction, wavelet augmentation, and pseudo-label refinement process.

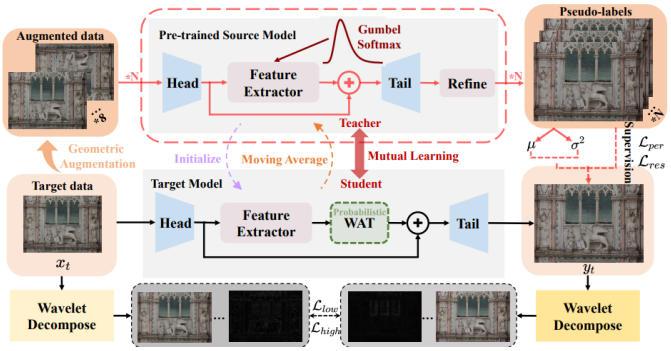


Fig. 3: Architecture of the SODA-SR framework. One target LR input image and its augmented versions are processed by the teacher model to generate refined pseudo-labels, which guide the student model's adaptation.

3) Wavelet Augmentation Transformer (WAT): WAT integrates wavelet transformations into the SR model to capture multi-scale and multi-frequency information:

- **Multi-Level Decomposition**: Decomposes feature maps into multiple frequency sub-bands using discrete wavelet transforms (DWT), capturing both global and local features.
- **Batch Augmentation Attention (BAA)**: Enhances features by attending over augmented versions across the batch dimension, increasing diversity and robustness.
- **Cross-Level Interaction**: Utilizes deformable attention mechanisms to aggregate information across different wavelet levels, enabling the model to learn complex dependencies.
- **Multi-Level Fusion**: Fuses the multi-frequency features to reconstruct the HR image, ensuring consistency and detail preservation.

4) Uncertainty-Aware Self-Training: To mitigate the impact of incorrect pseudo-labels, the method incorporates uncertainty estimation:

- **Knowledge Distillation**: Maintains a teacher model updated via exponential moving average (EMA) of the student model parameters, providing stable pseudo-labels.
- **Uncertainty Estimation**: Uses Gumbel-Softmax sampling to estimate the confidence in predictions, identifying unreliable regions.
- **Pseudo-Label Refinement**: Applies confidence thresholds to filter unreliable pseudo-labels, improving training stability and performance.

5) Regularization Losses: The framework introduces frequency-domain regularization to guide the adaptation process:

- **Low-Frequency Loss ($\mathcal{L}_{LF}$)**: Constrains the low-frequency components to match between the generated and real images, preserving overall image structure.

- **High-Frequency Adversarial Loss ($\mathcal{L}_{\text{HF}}$)**: Encourages the preservation of high-frequency details by adversarially training on high-frequency components, enhancing textures and edges.

The total loss function combines reconstruction loss, uncertainty-weighted self-training loss, and regularization losses:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{rec}} + \lambda_{\text{self}}\mathcal{L}_{\text{self}} + \lambda_{\text{LF}}\mathcal{L}_{\text{LF}} + \lambda_{\text{HF}}\mathcal{L}_{\text{HF}} \tag{11}$$

6) Implementation Details: The model architecture builds upon transformer-based designs, integrating wavelet transformations within the attention mechanisms. The teacher and student models share the same architecture but differ in parameter updates due to EMA.

Training involves initializing the student model with pretrained source weights. The model is then fine-tuned on the target domain using only unlabeled data, with the uncertainty-aware self-training guiding the adaptation process.

IV. Experimental Results

A. DASRGAN

1) Datasets: Experiments were conducted on the M3FD dataset [14], which contains paired LR-HR IR images with varying noise levels and resolutions. The dataset provides a realistic benchmark for evaluating IR image SR methods.

2) Evaluation Metrics: The performance of DASRGAN was evaluated using the following metrics:

- **Peak Signal-to-Noise Ratio (PSNR):** Quantifies reconstruction quality by measuring the similarity between the generated and ground truth images.
- **Structural Similarity Index Measure (SSIM):** Assesses perceptual similarity and structural fidelity of the reconstructed images.
- **Natural Image Quality Evaluator (NIQE):** Evaluates the perceptual realism of the generated images, with lower scores indicating better quality.

These metrics provide a comprehensive evaluation of both objective accuracy and perceptual quality.

3) Implementation Details: The model was implemented using PyTorch, and training was conducted on NVIDIA GPUs (HPC IIT Dharwad). Training on the M3FD dataset required approximately 48 hours for 100 epochs. The computational cost of DASRGAN was higher than baseline methods due to the alternating training phases for texture and noise-oriented adaptation. Input images were resized to 128×128, and standard data augmentation techniques (e.g., random cropping, flipping) were applied during training.

DASRGAN outperforms existing methods, achieving higher PSNR and SSIM values and lower NIQE scores, indicating better reconstruction quality and perceptual realism.

4) Qualitative Results: Figure 4 demonstrates the effectiveness of DASRGAN in reconstructing high-resolution IR images. Compared to other methods, DASRGAN achieves sharper textures and significantly reduces noise, particularly in regions with complex details or high degradation.

TABLE I: Quantitative comparison on the M3FD dataset.

Method	PSNR (dB)	SSIM	NIQE
Bicubic	24.12	0.678	5.67
SRCNN [4]	25.22	0.689	5.35
SRGAN [9]	26.38	0.731	4.92
ESRGAN [10]	27.41	0.754	4.68
DASRGAN (Ours)	**28.84**	**0.802**	**4.21**

(a) LR IR Image (b) DASRGAN SR Image

Fig. 4: Visual comparison of IR image SR using DASRGAN.

5) Quantitative Results: DASRGAN achieves state-of-the-art performance on the M3FD dataset compared to baseline methods, as shown in Table I. Improvements in PSNR and SSIM indicate superior reconstruction quality, while lower NIQE scores demonstrate enhanced perceptual realism. This highlights the effectiveness of domain adaptation in addressing the unique challenges of infrared image super-resolution.

6) Ablation Studies: Ablation studies were conducted to evaluate the contributions of individual components in DASRGAN:

- **Domain Adaptation Effectiveness:** Performance comparisons with and without visible-light domain adaptation showed significant improvements in PSNR and SSIM metrics.
- **Structural Ablation:** Incremental addition of modules (e.g., texture-oriented discriminator and noise adversarial loss) resulted in consistent performance gains.
- **Texture Prior Ablation:** Middle-level priors were found to be most effective for enhancing IR image details, outperforming shallow and deep priors.
- **Hyperparameter Tuning:** Optimal weights for noise and texture-oriented losses were identified as 0.1 and 1.0, respectively.

These experiments highlight the importance of each component in DASRGAN's architecture and training strategy.

B. Unsupervised Real-World Super-Resolution

1) Datasets: Experiments were conducted on the DIV2K [15] and Flickr2K [16] datasets, using unpaired LR and HR images. The NTIRE 2020 Real-World SR Challenge datasets [17] were also used for evaluation.

2) Evaluation Metrics: The performance of the proposed method was evaluated using the following metrics:

- **Peak Signal-to-Noise Ratio (PSNR):** Measures reconstruction quality by comparing the similarity between the reconstructed and ground truth images.

- **Structural Similarity Index Measure (SSIM):** Assesses structural fidelity and perceptual similarity.
- **Learned Perceptual Image Patch Similarity (LPIPS):** Evaluates perceptual quality, where lower values indicate better image quality.

These metrics ensure a balanced assessment of accuracy and perceptual realism.

3) Implementation Details: The model was trained with learning rates of 1×10^{-4} for the generator and 4×10^{-4} for the discriminators. Hyperparameters were set as $\lambda_{\text{adv}} = 0.01$, $\lambda_{\text{cyc}} = 10$, and $\lambda_{\text{id}} = 5$. Data augmentation included random cropping to 128×128 patches.

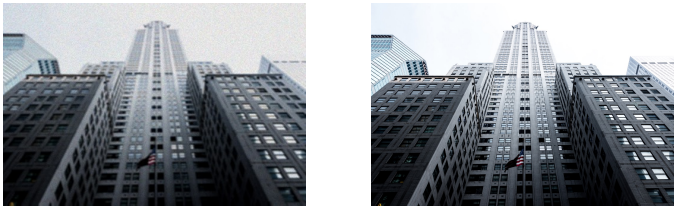
4) Quantitative Results: Table II presents the results on the DIV2K dataset, demonstrating the effectiveness of the proposed method.

TABLE II: Results on DIV2K Dataset.

Method	PSNR (dB)	SSIM	LPIPS
Bicubic	24.12	0.678	0.512
ZSSR [18]	22.17	0.472	0.412
Proposed Method	**23.39**	**0.537**	**0.342**

The proposed method achieves competitive results, improving upon baseline methods without requiring paired datasets.

5) Qualitative Results: Visual comparisons (Figure 5) show that the proposed method generates SR images with enhanced details and reduced artifacts.



(a) Real LR Image (b) SR Image (Proposed)

Fig. 5: Visual results on real-world images using the proposed method.

6) Ablation Studies (Summary): The ablation study highlights the contributions of individual components:

- **Feature Alignment:** Improved LPIPS and SSIM significantly by aligning feature distributions.
- **Cycle Consistency Loss:** Preserved content structure, enhancing PSNR and SSIM.

These components collectively contribute to the superior performance of the full model.

C. SODA-SR

1) Datasets and Metrics: Experiments were conducted on the DRealSR [19] dataset, a real-world SR benchmark with images captured by multiple DSLR cameras. Evaluation metrics included PSNR, SSIM, and LPIPS.

2) Implementation Details: The student model was initialized with pre-trained source weights. The model was trained using the Adam optimizer with learning rates of 1×10^{-4}. Hyperparameters were set as $\lambda_{\text{self}} = 1$, $\lambda_{\text{LF}} = 0.1$, and $\lambda_{\text{HF}} = 0.01$.

3) Quantitative Results: Table III summarizes the performance.

TABLE III: Quantitative Results on DRealSR Dataset.

Method	PSNR (dB)	SSIM	LPIPS
CinCGAN	30.17	0.821	0.391
DASR	30.86	0.828	0.372
DRN-Adapt	31.08	0.829	0.369
SODA-SR (Ours)	**31.22**	**0.832**	**0.344**

SODA-SR achieves state-of-the-art results, demonstrating its effectiveness in source-free domain adaptation.

4) Qualitative Results: Figure 6 illustrates the qualitative performance of SODA-SR on the DRealSR dataset. Compared to baseline methods, SODA-SR demonstrates superior capability in preserving fine textures, enhancing high-frequency details, and minimizing artifacts in real-world SR tasks.

(a) LR Image (b) SODA-SR SR Image

Fig. 6: Visual comparison of SR results using SODA-SR on real-world images.

These results highlight the effectiveness of wavelet-based augmentation and uncertainty-aware self-training in addressing the challenges of source-free domain adaptation for image SR.

5) Ablation Studies: Table IV shows the impact of key components.

TABLE IV: Ablation Study Results for SODA-SR.

Variant	PSNR (dB)	SSIM	LPIPS
Without WAT	30.64	0.822	0.365
Without Regularization	31.31	0.829	0.356
Full Model (Ours)	**31.41**	**0.832**	**0.344**

Results confirm the importance of WAT and regularization losses in improving performance.

V. Comparative Analysis

A. Discussion

DASRGAN excels in enhancing textures and suppressing noise in IR images through tailored discriminators and an alternating training strategy. However, it requires paired datasets,

TABLE V: Comparison of the Three Methods.

Aspect	DASRGAN	Unsupervised SR	SODA-SR
Domain	Infrared Imaging	Real-World Images	Synthetic to Real
Data Requirement	Paired LR-HR IR Images	Unpaired LR and HR Images	Unlabeled Target Data
Noise Handling	Excellent	Moderate	Good
Texture Enhancement	High	High	High
Generalization	Domain-Specific	Broad	Broad
Privacy Compliance	Not Addressed	Not Addressed	Yes
Computational Cost	Moderate	Moderate	Moderate

which may be difficult to obtain, and is specific to the IR domain.

The unsupervised SR method effectively bridges the domain gap without requiring paired datasets, making it practical for real-world applications. It generalizes well to unseen degradations but may not handle domain-specific noise as effectively as DASRGAN.

SODA-SR introduces a novel SFDA framework, leveraging wavelet-based augmentation and uncertainty-aware self-training to adapt to new domains without accessing source data, addressing privacy concerns. It demonstrates superior performance in adapting pre-trained models to target domains using only unlabeled data.

VI. Future Work

A. Multi-Domain Super-Resolution

Future research could focus on developing models capable of handling multiple domains simultaneously by integrating domain adaptation techniques that generalize across different data types, such as visible light, infrared, and medical images.

B. Adaptive Loss Functions

Designing adaptive loss functions that dynamically adjust weighting factors based on data characteristics could improve training stability and performance, particularly in scenarios with varying noise levels and degradations.

C. Real-Time Applications

Optimizing models for real-time SR applications is essential for deployment in resource-constrained environments like mobile devices or embedded systems. Techniques such as model pruning, quantization, and efficient network architectures could be explored.

D. Perceptual Quality Enhancement

Improving perceptual quality metrics (e.g., NIQE, LPIPS) to produce SR images that are not only quantitatively superior but also visually pleasing is crucial. Incorporating human perceptual models and conducting user studies could guide this enhancement.

E. Data Augmentation Strategies

Investigating advanced data augmentation techniques to simulate a wider range of real-world degradations could enhance model robustness. Methods such as generative modeling of degradations and domain randomization may be beneficial.

VII. Conclusion

This paper provides a comprehensive analysis of three advanced SR methods that leverage domain adaptation to address challenges in different domains. DASRGAN demonstrates the effectiveness of targeted domain adaptation in enhancing texture and suppressing noise in IR images. The unsupervised SR framework shows how feature distribution alignment can bridge the domain gap without requiring paired datasets. SODA-SR sets a new benchmark for source-free domain adaptation in SR tasks by introducing wavelet augmentation and uncertainty-aware self-training.

By exploring these approaches, we highlight the significant impact of domain adaptation in advancing SR performance across diverse applications. Future research can build upon these insights to develop more versatile and robust SR models, ultimately contributing to the broader field of image restoration and enhancement.

References

[1] Y. Huang, T. Miyazaki, X. Liu, and S. Omachi, "Target-Oriented Domain Adaptation for Infrared Image Super-Resolution," *arXiv preprint arXiv:2304.05601*, 2023. Available: https://github.com/yongsongH/DASRGAN

[2] W. Wang, H. Zhang, Z. Yuan, and C. Wang, "Unsupervised Real-World Super-Resolution: A Domain Adaptation Perspective," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2021, pp. 3645-3654.

[3] Y. Ai, X. Zhou, H. Huang, et al., "Uncertainty-Aware Source-Free Adaptive Image Super-Resolution with Wavelet Augmentation Transformer," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024.

[4] C. Dong, C. C. Loy, K. He, and X. Tang, "Learning a Deep Convolutional Network for Image Super-Resolution," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2014, pp. 184-199.

[5] J. Kim, J. K. Lee, and K. M. Lee, "Accurate Image Super-Resolution Using Very Deep Convolutional Networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 1646-1654.

[6] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee, "Enhanced Deep Residual Networks for Single Image Super-Resolution," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2017, pp. 136-144.

[7] Y. Zhang et al., "Image Super-Resolution Using Very Deep Residual Channel Attention Networks," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 286-301.

[8] I. Goodfellow et al., "Generative Adversarial Nets," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2014, pp. 2672-2680.

[9] C. Ledig et al., "Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 4681-4690.

[10] X. Wang et al., "ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks," in *Proceedings of the European Conference on Computer Vision Workshops (ECCVW)*, 2018, pp. 63-79.

[11] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 2223-2232.

[12] K. Zhang et al., "Designing a Practical Degradation Model for Deep Blind Image Super-Resolution," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021, pp. 4791-4800.

[13] L. Wang et al., "A Survey of Infrared Image Super-Resolution," *Infrared Physics & Technology*, vol. 104, p. 103127, 2020.

[14] X. Liu et al., "M3FD: Multi-Modal Multi-Scale Multi-Spectral Image Fusion Dataset," *IEEE Transactions on Image Processing*, vol. 29, pp. 4789-4802, 2020.

[15] E. Agustsson and R. Timofte, "NTIRE 2017 Challenge on Single Image Super-Resolution: Dataset and Study," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2017, pp. 126-135.

[16] Y. Timofte et al., "NTIRE 2017 Challenge on Single Image Super-Resolution: Methods and Results," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2017, pp. 114-125.

[17] J. Gu et al., "NTIRE 2020 Challenge on Real-World Image Super-Resolution: Methods and Results," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2020, pp. 494-495.

[18] A. Shocher, N. Cohen, and M. Irani, "Zero-Shot Super-Resolution Using Deep Internal Learning," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 3118-3126.

[19] K. Wei et al., "DRealSR: A Benchmark for Real World Image Super-Resolution," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021.