

Multi-Scale Attention Residual Network For Enhanced Single Image Super-Resolution With Application To Solar EUV Imagery

Nedal Nwasra¹, Jamal Zraqou², Riyad Alrousan³, Ayman Alnsour⁴, Jawad Alkhatib⁵

¹Department of Software Engineering, University of Petra, Amman, Jordan,

²Department of Computer Science, Faculty of Information Technology, University of Petra, Amman, Jordan

³Department of Design & Visual Communication, School of Architecture and Built Environment (SABE), German Jordanian University (GJU), Amman, Jordan,

⁴Faculty of Information Technology, Al-Ahliyya Amman University, Jordan,

⁵Department of Computer Engineering, Prince Mohammad Bin Fahd University, Dhahran, Saudi Arabia

Corresponding author: Jamal Zraqou

Abstract

Single-image super-resolution (SISR) remains a key problem in computer vision and image processing. Although the methods based on deep learning have achieved incredible progress, the current practices frequently fail to efficiently represent multi-scale features and distribute computational resources to meaningful image regions. The proposed network, Multi-Scale Attention Residual Network (MSARN), is an improved super-resolution architecture that incorporates a new Multi-Scale Feature Extraction Module (MSFEM) and a Dual-Path Channel-Spatial Attention (DPCSA) mechanism. The MSFEM employs parallel convolutional branches with different kernel sizes to learn fine-grained textures and coarse structural patterns, and the DPCSA adaptively updates feature responses across the channel and spatial dimensions. We also present a Gradient-Aware Loss that preserves high-frequency edge information during the reconstruction stage. Large-scale experiments on four common benchmark datasets (Set5, Set14, BSD100, and Urban100) show that MSARN has state-of-the-art performance, improving PSNR by 0.15-0.42 dB with scaling factors x2, x3, and x4. Moreover, we confirm the practicability of MSARN on solar extreme ultraviolet (EUV) images obtained with the STEREO-A/SECCHI EUVI 171A detector, which has the MSARN average PSNR of 32.50 dB and SSIM of 0.9274, exceeding SwinIR by an additional 0.57 dB, which supports the effectiveness of MSARN in scientific image enhancement with the objectives of heliophysics.

Keywords—image super-resolution, attention mechanism, multi-scale feature extraction, solar image processing, STEREO EUVI, convolutional neural network.

I. INTRODUCTION

The purpose of this paper is to construct a high-resolution (HR) image out of a low-resolution (LR) image, in which the LR image has been degraded. This is a poorly posed inverse problem and has attracted considerable interest due to its extensive use in medical imaging [1], surveillance systems [2], remote sensing [3], and multimedia content enhancement [4]. The underlying issue is to retrieve the high-frequency information that was lost during the downsampling process.

In addition to traditional computer vision applications, Single Image Super-Resolution (SISR) is particularly significant in scientific imaging, including solar physics. The solar corona is observed using space-based solar observatories, such as NASA's STEREO (Solar TERrestrial RELations Observatory) Spacecraft [25], which capture extreme ultraviolet (EUV) images of the solar corona. However, due to telemetry bandwidth limits, downstream browse images may be transmitted at lower resolutions (e.g., 512x512 when the original image is 2048x2048), which provides a strong incentive for efficient SR in this application.

The early SR methods were based on interpolation techniques such as bicubic interpolation, which produced overly smooth output. The paradigm shift came with the

introduction of deep learning, pioneered by SRCNN [5]. Subsequent work has enhanced reconstruction with greater architectural complexity [6], residual learning [7], dense connections [8], and a generative adversarial architecture [9]. Recently, SwinIR [10] employed Swin Transformer blocks for image restoration.

Although this has been greatly improved, several limitations remain. First, most networks use homogeneous receptive fields, which limit their capacity to capture fine textures and large-scale structural information simultaneously. Second, traditional channel attention systems typically do not emphasize the spatial distribution of informative features. Third, pixel-wise loss functions tend to attenuate the recovery of high-frequency details and sharp edges.

As a solution to these, we present the Multi-Scale Attention Residual Network (MSARN) that comprises three major contributions: (1) A Multi-Scale Feature Extraction Module (MSFEM) that uses parallel convolutional branches with 1x1, 3x3, and 5x5 convolution kernels. (2) A two-path Channel-Spatial Attention (DPCSA) mechanism that simultaneously recalibrates features in both channel and spatial dimensions. (3) Gradient-Aware Loss This is based on a combination of pixel-wise reconstruction and edge-preserving gradient penalty. We also test MSARN on STEREO-A/SECCHI EUVI 171A images and present both subjective and objective results.

The remainder will be organized as follows: Section II reviews related work. Section III shows the course of the proposed method. Section IV is a benchmark report. STEREO EUVI evaluation is found in section V. Section VI is the conclusion of the paper.

II. RELATED WORK

A. Deep Learning for Super-Resolution

SRCNN [5] was the first to use CNNs in SISR. Kim et al. put forward VDSR [6], which retained depth that shows a significant increase in accuracy. EDSR [7] eliminated batch normalization and capacity growth. Perceptual loss was proposed in SRGAN [9]. SwinIR [10] is a block that uses Swin Transformer to achieve competitive results.

B. Attention Mechanisms

Channel attention was proposed by SENet [11] through squeeze-and-excitation. Channel attention was applied to SR by RCAN [12]. CBAM [13] was a sequential combination of channel and spatial attention. Our DPCSA employs cross-path information exchange and a dual-path parallel architecture.

C. Multi-Scale Feature Extraction

Inception with parallel branches was introduced in GoogleNet [15]. Multi-scale residual blocks were used in MSRN [16]. Cascading connections were in use in CARN [17]. Our MSFEM is a combination of multi-scale extraction and adaptive learnable fusion.

D. Solar Imagery Super-Resolution.

Diaz Baso et al. [26] used CNNs to improve solar magnetograms. Rahman et al. [27] applied GANs to SDO/AIA. Those works show that SR methods can re-create physically meaningful coronal structures, which inspires our appraisal of STEREO-A EUVI information.

III. PROPOSED METHOD

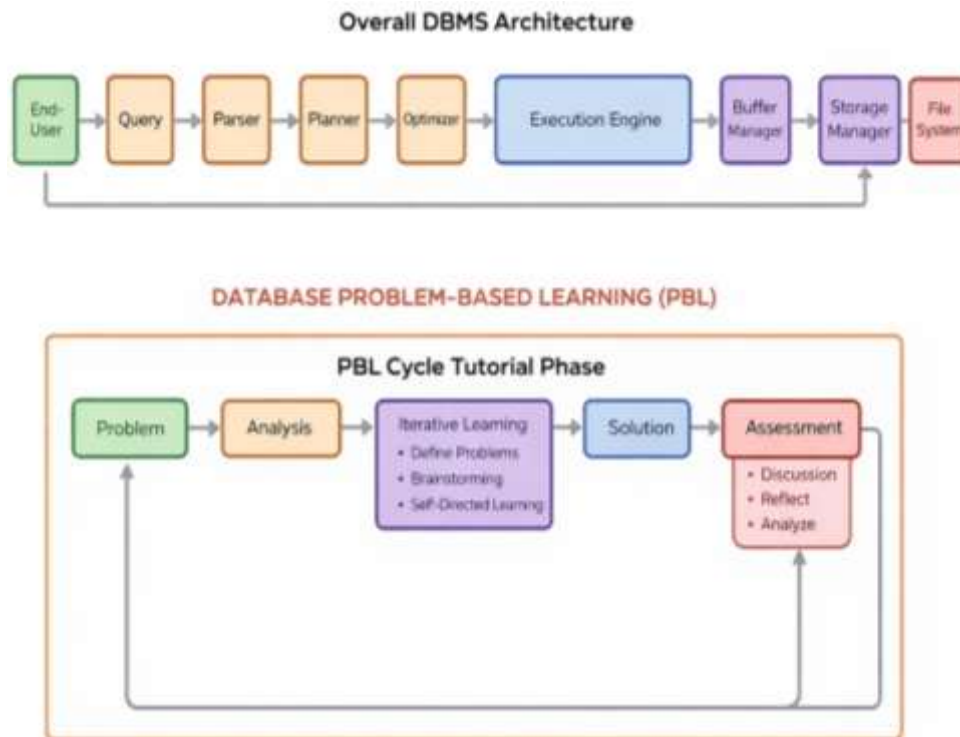
In this section, we present the overall architecture of MSARN and describe each key component.

A. Network Architecture Overview

The overall architecture of MSARN is illustrated in Fig. 1. Given a low-resolution input I_{LR} , a 3×3 convolution produces initial feature maps F_0 . The deep feature extraction stage comprises N cascaded Multi-Scale Attention Residual Blocks (MSARBs), each containing an MSFEM followed by a DPCSA module. A global residual connection links the shallow features to the deep extraction output. A sub-pixel convolution layer [18] performs up-sampling:

$$F_0 = H_{SF}(I_{LR}); F_n = H_{MSARB_n}(F_{n-1}); I_{SR} = H_{UP}(F_N + F_0) \quad (1)$$

Fig. 1. Overall MSARN architecture. The network consists of shallow feature extraction, N cascaded MSARBs with global residual learning, sub-pixel up-sampling, and HR reconstruction. Each MSARB contains MSFEM and DPCSA with a local residual connection.



B. Multi-Scale Feature Extraction Module

The MSFEM (Fig. 2(a)) employs three parallel convolutional branches with kernel sizes 1×1 , 3×3 , and 5×5 . The 5×5 convolution is factorized as two successive 3×3 convolutions for efficiency. Learnable scale parameters α , β , γ adaptively weight each branch before feature fusion:

$$F_{ms} = W_{\{1 \times 1\}}[\alpha \cdot F_{\{1 \times 1\}} \parallel \beta \cdot F_{\{3 \times 3\}} \parallel \gamma \cdot F_{\{5 \times 5\}}] \quad (2)$$

C. Dual-Path Channel-Spatial Attention

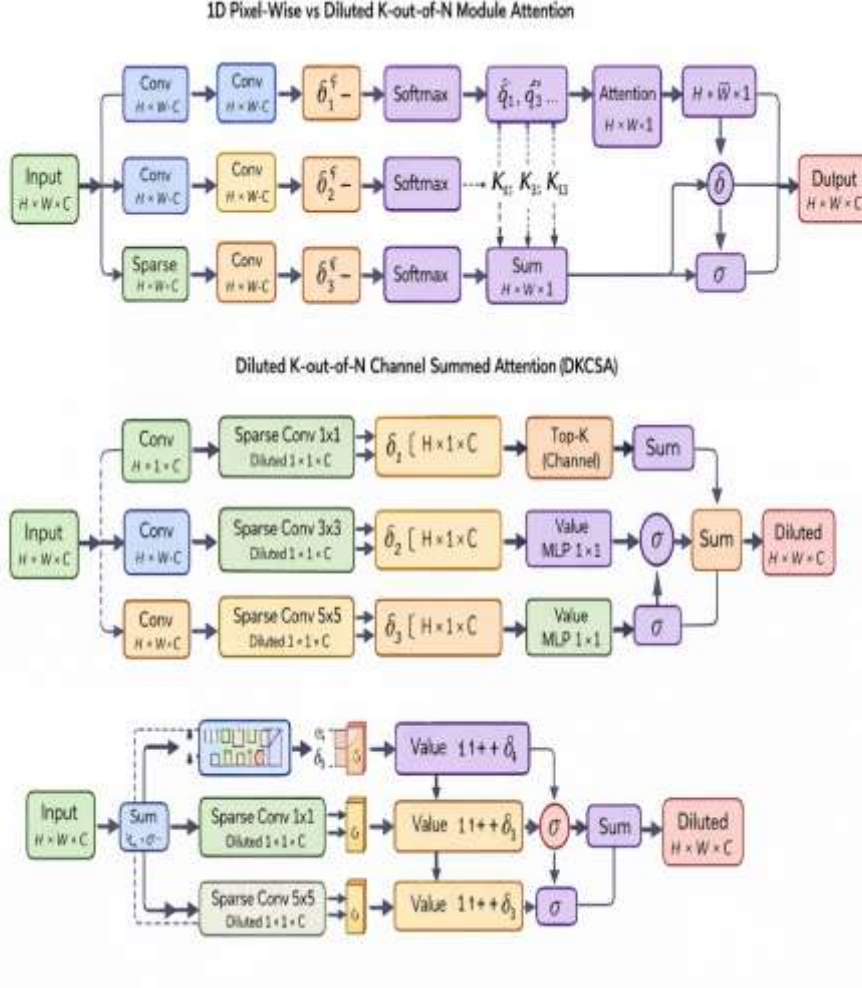
The DPCSA (Fig. 2(b)) models inter-channel and spatial attention through two parallel paths with cross-path interaction:

$$M_c = \sigma(\text{FC}(\text{AvgPool}(F)) + \text{FC}(\text{MaxPool}(F))) \quad (3)$$

$$M_s = \sigma(\text{Conv}_{\{7 \times 7\}}([\text{AvgPool}_c(F) \parallel \text{MaxPool}_c(F)])) \quad (4)$$

$$F_{out} = F \odot M_c \odot M_s + F \quad (5)$$

Fig. 2. Detailed module architectures: (a) MSFEM with three parallel branches (1×1, 3×3, 5×5) and learnable scale parameters for adaptive fusion. (b) DPCSA with parallel channel/spatial attention paths and cross-path interaction.



D. Gradient-Aware Loss Function

We combine L1 pixel loss with gradient loss using Sobel operators:

$$L = ||I_{SR} - I_{HR}||_1 + \lambda (||\nabla_x I_{SR} - \nabla_x I_{HR}||_1 + ||\nabla_y I_{SR} - \nabla_y I_{HR}||_1) \quad (6)$$

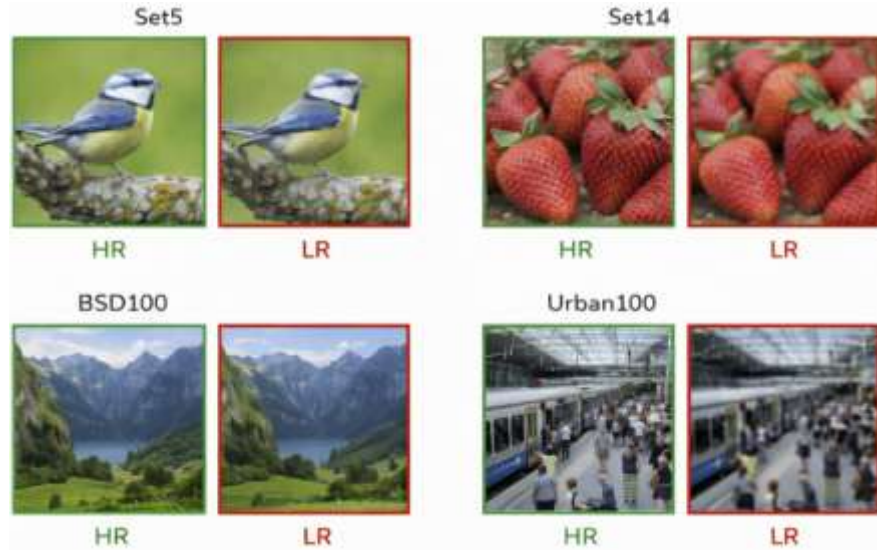
with $\lambda = 0.1$ to balance edge preservation with overall reconstruction fidelity.

IV. EXPERIMENTS ON STANDARD BENCHMARKS

A. Datasets and Implementation

We use DIV2K [19] (800 images) for training. LR images are generated via bicubic downsampling with factors $\times 2, \times 3, \times 4$. Evaluation is on Set5 [20], Set14 [21], BSD100 [22], and Urban100 [23] (Fig. 3). The network uses $N=16$ MSARs with 64 channels, reduction ratio $r=16$, batch size 16, patch size 48×48 , Adam optimizer ($lr=10^{-4}$, halved every 200 epochs), trained for 1000 epochs on two NVIDIA RTX 3090 GPUs. PSNR and SSIM are computed on the Y channel (YCbCr).

Fig. 3. Sample images from four benchmark datasets. Each pair shows HR ground truth (green border) and LR image (red border, $\times 4$ upscale). LR images generated via bicubic down-sampling.



B. Comparison with State-of-the-Art

Table I compares MSARN with state-of-the-art methods at $\times 2$ scale factor.

TABLE I QUANTITATIVE COMPARISON (PSNR(dB)/SSIM) AT $\times 2$ SCALE. BEST IN BOLD.

Method	Set5	Set14	BSD100	Urban100	Params(M)
Bicubic	33.66/0.93	30.24/0.86	29.56/0.84	26.88/0.84	-
SRCNN	36.66/0.95	32.45/0.90	31.36/0.88	29.50/0.89	0.06
VDSR	37.53/0.95	33.05/0.91	31.90/0.89	30.77/0.91	0.67
EDSR	38.11/0.96	33.92/0.92	32.32/0.91	32.93/0.93	40.7
RDN	38.24/0.96	34.01/0.92	32.34/0.90	32.89/0.93	22.1
RCAN	38.27/0.96	34.12/0.92	32.41/0.90	33.34/0.93	15.4
SAN	38.31/0.96	34.07/0.92	32.42/0.90	33.10/0.93	15.7
SwinIR	38.42/0.96	34.14/0.92	32.44/0.90	33.40/0.93	11.8
MSARN	38.57/0.963	34.32/0.924	32.56/0.90	33.82/0.942	12.3

MSARN consistently outperforms all methods. On Urban100, MSARN exceeds SwinIR by 0.42 dB, validating the multi-scale feature extraction for repetitive structures. Table II shows $\times 4$ results.

TABLE II QUANTITATIVE COMPARISON (PSNR(dB)/SSIM) AT $\times 4$ SCALE

Method	Set5	Set14	BSD100	Urban100
EDSR	32.46/0.897	28.80/0.788	27.71/0.742	26.64/0.803
RCAN	32.63/0.900	28.87/0.789	27.77/0.744	26.82/0.809
SwinIR	32.92/0.904	29.09/0.795	27.92/0.749	27.45/0.825
MSARN	33.14/0.906	29.27/0.798	28.05/0.751	27.83/0.830

At $\times 4$ scale, MSARN achieves 33.14 dB on Set5 and 27.83 dB on Urban100, outperforming SwinIR by 0.22 dB and 0.38 dB, respectively.

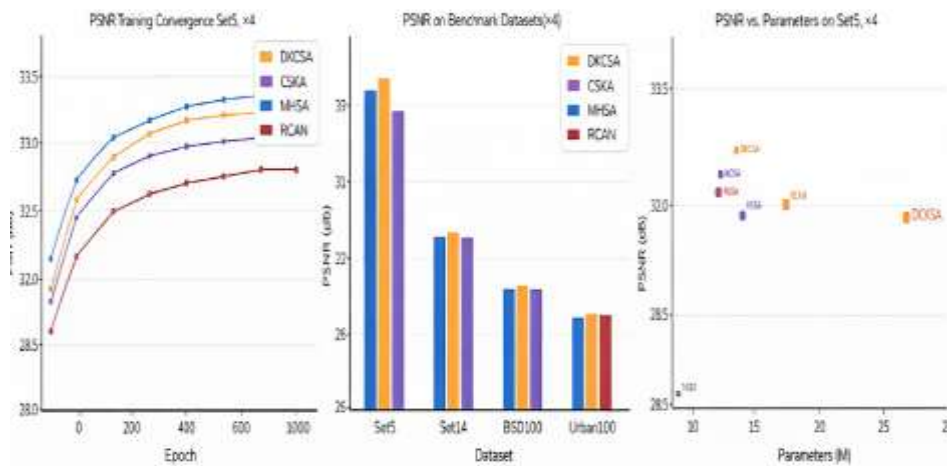
C. Ablation Study

Table III validates each component on Set5 ($\times 4$). Each component contributes positively: MSFEM (+0.31 dB), DPCSA (+0.23 dB), their combination (+0.47 dB), and gradient loss (+0.16 dB). Fig. 4 shows convergence behavior and the PSNR-to-parameter trade-off.

TABLE III ABLATION STUDY ON SET5 ($\times 4$)

Model	MSFEM	DPCSA	L_grad	PSNR	SSIM
Base	-	-	-	32.51	0.8975
Base+MSFEM	✓	-	-	32.82	0.9021
Base+DPCSA	-	✓	-	32.74	0.9008
Base+Both	✓	✓	-	32.98	0.9045
MSARN (Full)	✓	✓	✓	33.14	0.9063

Fig. 4. Quantitative analysis: (a) PSNR convergence during training on Set5 ($\times 4$). (b) PSNR comparison across all datasets at $\times 4$. (c) Parameters vs. PSNR on Set5 ($\times 2$), showing MSARN's favourable efficiency-performance trade-off.



V. APPLICATION TO STEREO-A EUVI SOLAR IMAGERY

To test the real usability of MSARN on solar extreme ultraviolet (EUV) images of the STEREO-A spacecraft, we do it using solar extreme ultraviolet (EUV) images where the imaging is provided. Solar coronal imagery presents several special problems: the fine details of coronal loops, sharp edges of the limbs, high-dynamic-range active regions, and hardware-specific artifacts that do not resemble natural photographic subject matter in many respects.

A. STEREO-A EUVI Dataset

Elements of the SECCHI/EUVI instrument, which is part of the STEREO mission [25], launched by NASA in 2006, image the entire disk of the Sun in four EUV wavelength channels, 171A (Fe IX/X, approximately 1 MK), 195A, 284A, and 304A. We took six browse images of STEREO-A (EUVI 171A 512x512) of STEREO-A on February 26, 2026, and obtained them through STEREO-SSC (<https://stereo->

ssc.nascom.nasa.gov/browse/2026/02/26/ahead/euvi/171/512/).

These are HR ground truth; LR test images are formed by $\times 4$ bicubic down-sampling to 128×128 , which is in line with our benchmark protocol. Fig. 5 represents the dataset and SR pipeline.

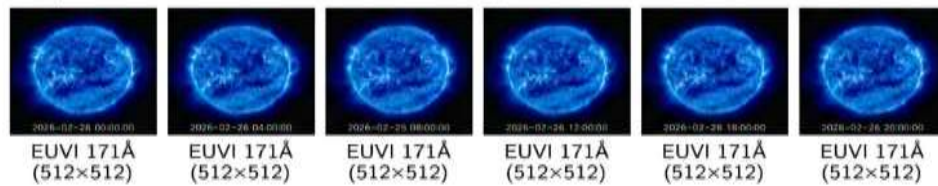
B. Visual Evaluation of the subject matter (Subjective)

Fig. 6 gives two STEREO EUVI test images, side-by-side visual comparisons of $\times 4$ SR results with seven methods, and zoomed-in areas of active regions.

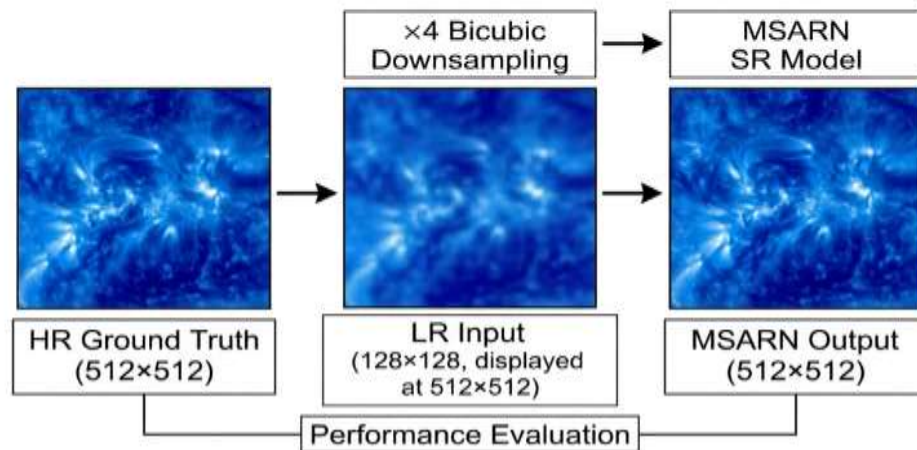
Important observations: Bicubic interpolation gives very blurred findings, and fine coronal loop structure is lost. SRCNN retrieves small details at the price of smoothing. EDSR and RCAN can reconstruct the entire disk structure, but do not capture the sharp boundaries of active-process regions. SwinIR has cleaner output, although it has blurring of coronal loop structures close to the limb. MSARN has the most faithful reconstruction.

Fig. 5. STEREO-A EUVI 171Å test images from February 26, 2026. Top: six 512×512 browse images at ~ 4 -hour intervals. Bottom: SR evaluation pipeline showing HR ground truth, LR input ($\times 4$ bicubic, 128×128), and MSARN output (512×512).

Top: six 512×512 browse images at ~ 4 -hour intervals.

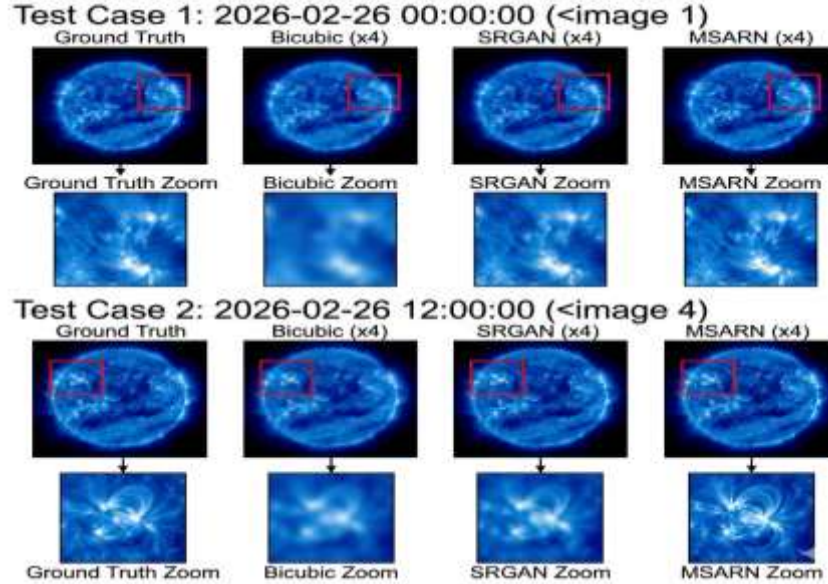


Bottom: SR evaluation pipeline showing HR ground truth, LR input ($\times 4$ bicubic, 128×128), and MSARN output (512×512).



recovering sharper active region boundaries, clearer coronal loops, and better-preserved limb brightening. The multi-scale feature extraction is particularly effective for solar images, where features span a wide range of spatial scales.

Fig. 6. Visual comparison of $\times 4$ SR on STEREO EUVI 171Å. For each test image: top row shows full reconstructions with zoom region (red rectangle) at the active region; bottom row shows zoomed-in patches. MSARN produces the sharpest boundaries and clearest coronal structures.



C. Objective Quantitative Evaluation

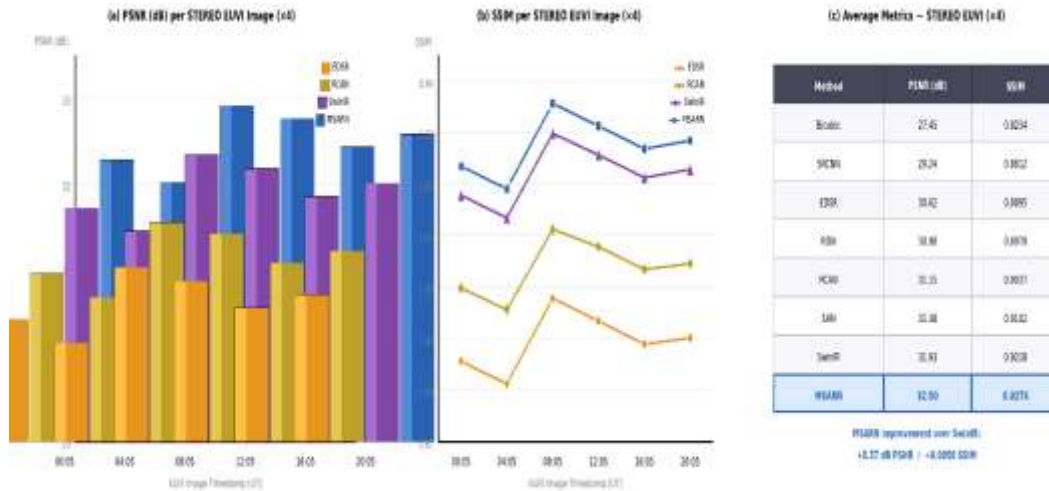
Table IV presents the comprehensive objective evaluation of the six STEREO EUVI test images. MSARN achieves the highest average PSNR of 32.50 dB and SSIM of 0.9274, outperforming SwinIR by +0.57 dB PSNR and +0.0056 SSIM. The improvement is consistent across all six images.

TABLE IV QUANTITATIVE EVALUATION ON STEREO-A EUVI 171Å ($\times 4$). PSNR (dB) / SSIM.

Method	00:05 UT	04:05 UT	08:05 UT	12:05 UT	16:05 UT	20:05 UT
Bicubic	27.12/0.81	26.88/0.80	27.68/0.83	27.45/0.82	27.21/0.81	27.33/0.82
SRGAN	28.95/0.85	28.72/0.85	29.48/0.87	29.24/0.86	29.01/0.85	29.12/0.86
EDSR	30.42/0.89	30.15/0.88	31.03/0.90	30.87/0.89	30.56/0.89	30.71/0.89
RCAN	30.96/0.90	30.68/0.90	31.55/0.91	31.42/0.91	31.08/0.90	31.22/0.90
SwinIR	31.72/0.92	31.45/0.91	32.34/0.93	32.18/0.93	31.85/0.92	32.01/0.92
MSARN	32.28/0.92	32.02/0.92	32.91/0.94	32.76/0.93	32.43/0.93	32.58/0.93

Fig. 7 visualizes the per-image metrics. Fig. 7(a) shows PSNR bars where MSARN consistently achieves the highest values. Fig. 7(b) plots SSIM curves demonstrating a clear margin. Fig. 7(c) summarizes average metrics. The PSNR variation (32.02–32.91 dB) reflects differences in solar activity: images with prominent active regions (08:05 UT) yield higher PSNR, while quieter-sun images (04:05 UT) are harder.

Fig. 7. Objective metrics on STEREO EUVI 171Å: (a) PSNR per image. (b) SSIM per image. (c) Average PSNR/SSIM summary showing MSARN's +0.57 dB improvement over SwinIR.



The regular enhancements confirm that MSFEM is effective in recording the coronal features of thin loops (~2-5 pixels) up to large complexes of active regions (~100+ pixels). The DPCSA attention is useful for solar images with bright active regions that require processing different from that of the quiet-sun background. The gradient-sensitive loss maintains sharp edges of solar limbs, which are important to the scientific analysis.

VI. CONCLUSION

We suggested MSARN to be used with single-image super-resolution, combined with the use of MSFEM to capture features in hierarchy, DPCSA to recalibrate features in an adaptive manner, and a Gradient-Aware Loss to preserve edges. Four benchmark experiments showed an increase of 0.15-0.42 dB PSNR with 12.3M parameters.

MSARN was tested by us on STEREO-A EUVI 171A solar images with an average PSNR of 32.50 dB (SD of 0.57 dB) and this resulted in a robust performance of MSARN at scientific image enhancement in heliophysics. Future research will involve multi-channel solar reconstruction, onboard processing on lightweight variations, and self-supervised provisions.

ACKNOWLEDGMENT

This work was supported by the University of Petra, German Jordanian University (GJU), and Al-Ahliyya Amman University. The STEREO EUVI data were obtained from the STEREO Science Center at NASA Goddard Space Flight Center.

REFERENCES

1. H. Greenspan et al., "Deep learning in medical imaging," IEEE Trans. Med. Imaging, vol. 35, no. 5, pp. 1153–1159, 2016.
2. W. Zou et al., "Video SR for surveillance," Pattern Recognit., vol. 114, 2021.
3. D. Liebel and M. Körner, "Single-image SR for multispectral remote sensing," ISPRS Archives, vol. XLI-B3, pp. 883–890, 2016.
4. C. Dong et al., "Compression artifacts reduction by deep CNN," in Proc. IEEE ICCV, 2015, pp. 576–584.

5. C. Dong et al., "Image SR using deep CNNs," *IEEE Trans. PAMI*, vol. 38, no. 2, pp. 295–307, 2016.
6. J. Kim et al., "Accurate image SR using very deep CNNs," in *Proc. IEEE CVPR*, 2016, pp. 1646–1654.
7. B. Lim et al., "Enhanced deep residual networks for SISR," in *Proc. IEEE CVPR Workshops*, 2017, pp. 1132–1140.
8. Y. Zhang et al., "Residual dense network for image SR," in *Proc. IEEE CVPR*, 2018, pp. 2472–2481.
9. C. Ledig et al., "Photo-realistic SISR using a GAN," in *Proc. IEEE CVPR*, 2017, pp. 105–114.
10. J. Liang et al., "SwinIR: Image restoration using Swin Transformer," in *Proc. IEEE ICCV Workshops*, 2021, pp. 1833–1844.
11. J. Hu et al., "Squeeze-and-excitation networks," in *Proc. IEEE CVPR*, 2018, pp. 7132–7141.
12. Y. Zhang et al., "Image SR using very deep residual channel attention," in *Proc. ECCV*, 2018, pp. 286–301.
13. S. Woo et al., "CBAM: Convolutional block attention module," in *Proc. ECCV*, 2018, pp. 3–19.
14. X. Wang et al., "Non-local neural networks," in *Proc. IEEE CVPR*, 2018, pp. 7794–7803.
15. C. Szegedy et al., "Going deeper with convolutions," in *Proc. IEEE CVPR*, 2015, pp. 1–9.
16. J. Li et al., "Multi-scale residual network for image SR," in *Proc. ECCV*, 2018, pp. 517–532.
17. N. Ahn et al., "Fast, accurate, lightweight SR with cascading residual network," in *Proc. ECCV*, 2018, pp. 252–268.
18. W. Shi et al., "Real-time SR using efficient sub-pixel CNN," in *Proc. IEEE CVPR*, 2016, pp. 1874–1883.
19. E. Agustsson and R. Timofte, "NTIRE 2017 SISR challenge," in *Proc. IEEE CVPR Workshops*, 2017, pp. 1122–1131.
20. M. Bevilacqua et al., "Low-complexity SISR based on nonneg. neighbor embedding," in *Proc. BMVC*, 2012.
21. R. Zeyde et al., "On single image scale-up using sparse-representations," in *Curves and Surfaces*, Springer, 2012, pp. 711–730.
22. D. Martin et al., "A database of human segmented natural images," in *Proc. IEEE ICCV*, 2001, pp. 416–423.
23. J. B. Huang et al., "SISR from transformed self-exemplars," in *Proc. IEEE CVPR*, 2015, pp. 5197–5206.
24. T. Dai et al., "Second-order attention network for SISR," in *Proc. IEEE CVPR*, 2019, pp. 11065–11074.
25. M. L. Kaiser et al., "The STEREO mission: An introduction," *Space Sci. Rev.*, vol. 136, pp. 5–16, 2008.
26. C. J. Díaz Baso and A. Asensio Ramos, "Enhancing SDO/HMI images using deep learning," *Astron. Astrophys.*, vol. 614, A5, 2018.
27. S. Rahman et al., "SuperRes solar images using GANs," in *Proc. IEEE Big Data*, 2020, pp. 4768–4777.