DDIFNet: A Dual-Domain Interactive Fusion Network for Image Restoration

Ying Yang, Shunpeng Xiu, Tao Jiang, Yelin Xia, Xiaohui Kou, Hanning Chen

Abstract— **Image restoration in real-world scenarios, especially in professional fields such as footprint analysis that demand strict detail fidelity, poses significant challenges to the design of network architectures. Most existing methods are tailored for general natural images; when professional images with specific structural textures and complex noise are processed, an optimal balance between denoising and detail preservation is often difficult to achieve. To address this challenge, a novel architecture named Dual-Domain Interactive Fusion Network (DDIFNet) is proposed in this paper, which aims to solve image restoration problems in specific domains.**

The core of DDIFNet is a novel Dual-Domain Fusion Module (DDFM), which is designed to process features from the spatial domain and the frequency domain (obtained via Fast Fourier Transform) in parallel. A key design of DDFM is an interactive fusion mechanism: features in the spatial domain (e.g., textures, edges) can guide the filtering process in the frequency domain, while features in the frequency domain (e.g., the spectral distribution of noise) can inversely modulate the feature responses in the spatial domain. This bidirectional interaction mechanism allows the network to adaptively integrate cross-domain information based on local image content.

To evaluate the effectiveness of DDIFNet, core experiments were conducted on a challenging self-constructed forensic footprint image dataset. Experimental results demonstrate that DDIFNet outperforms current state-of-the-art methods (including SwinIR) significantly in restoring key forensic features corrupted by severe noise. Furthermore, to verify its generalization capability, the network was tested on the general denoising benchmark SIDD, and competitive results are achieved. These results prove that the proposed architecture not only specializes in solving specific problems but also maintains good generality, which validates the advancement and practical value of its design.

Index Terms—**Image restoration, Footprint analysis, Dual-domain interactive fusion network, Fast Fourier trans-form, Denoising**

I. INTRODUCTION

Digital images have become ubiquitous in modern technology and daily life, yet their quality is often severely compromised by noise. Noise not only degrades the visual experience of images but also impairs the accuracy of subsequent advanced visual analysis—a problem that is particularly prominent in professional fields requiring strict attention to image details, such as footprint analysis and industrial non-destructive testing [29]. For instance, in crime scene investigations, footprint images captured under low-light conditions or from non-ideal surfaces are often heavily contaminated by noise, which directly under-mines the effective identification of critical forensic features (e.g., sole wear patterns). Thus, image denoising, as a classic problem in the field of image processing, remains a research focus to date [30]. Recent review papers have noted that deep learning has achieved remarkable progress in efficient image denoising, and emphasize that future research directions include lightweight net-work design and few shot adaptation [1]. Additionally, another review focusing on spatial frequency domain integration strategies proposes a novel approach that combines the transform do-main with self-attention networks [2]. Meanwhile, in addressing periodic and real-world noise, neural network strategies based on frequency do-main transformation have also demonstrated excellent SSE performance [40].

In recent years, deep learning methods represent-ed by Convolutional Neural Networks (CNNs) have made significant advancements in the field of image de-noising [23]. Architectures such as DnCNN [35] and UNet [29] have demonstrated competitive perfor-mance in general natural image denoising tasks [31]. However, most CNN architectures primarily operate in the spatial domain, and their inherent local receptive fields may limit their ability to capture the global distribution characteristics of noise. Although Trans-former-based models [24] have introduced global dependencies via self-attention mechanisms, the uni-form processing strategies of these general-purpose models may still lead to over-smoothing of critical fine-grained features when handling images with highly specialized textures (e.g., shoe sole patterns) and specialized noise patterns. To address this issue, se-veral studies have attempted to fuse CNNs with Transformers—for example, hybrid architectures that integrate CNN features and Transformer attention me-chanisms have been developed for detail preservation [36], and networks combining self-attention with spatial-frequency fusion have also achieved breakthroughs [5].

An ideal denoising network should be cap-able of optimization for specific problem do-mains. The characteristics of real-world noise are the result of the combined effects of the spatial and frequency domains [27]. For images such as forensic footprints, their sole patterns often exhibit periodicity, while wear marks manifest as unstructured features—and this complexity exhibits distinct representations in the spatial and frequency domains.

Manuscript received October 24, 2025

Ying Yang, Information Support Detachment, Binhai Branch, Tianjin Public Security Bureau, Tianjin 300456, China
Shunpeng Xiu,School of Software, Tiangong University, Tianjin 300387, China
Tao Jiang, Xinjiang Power Transmission and Distribution Co., Ltd., Urumqi, Xinjiang 830011, China
Yelin Xia,College of Artificial Intelligence, Tianjin University of Science and Technology, Tianjin 300457, China
Xiaohui Kou,School of Computer Science and Technology, Tiangong University, Tianjin 300387, China
Hanning Chen,School of Computer Science and Technology, Tiangong University, Tianjin 300387, China

Existing methods either focus on the spatial domain or treat frequency-domain process-ing as an auxiliary tool [15], and still lack deep, dyna-mic interaction mechanisms between the two domains to address such challenges. In response to this need, re-searchers have proposed frequency-domain hybrid Transformer architectures, which integrate convolution and global modeling at the spectral level [5]; addi-tionally, light-weight attention-fusion models have achieved significant improvements in handling Gauss-ian and real-world noise [6].

To fill this research gap, recent researchers have begun to explore deep fusion mechanisms be-tween the spatial and frequency domains. For example, the Multi-scale Adaptive Dual-domain Network (MADNet), proposed in 2025, enables dynamic interaction between spatial and frequency domain information by introducing Adaptive Spatial-Frequency Learning Units (ASFLUs) [7]. This architecture utilizes learned masks to separate high-frequency and low-frequency information, and combines the multi scale characte-ristics of image pyramids—effectively enhancing the ability to distinguish between global and local noise features and thereby significantly improving denoising performance [7]. Another study from 2024, focused on periodic noise in infrared scanning images, proposed a method that converts 2D images into 1D signals—either applying neural networks to predict noise or directly modeling noise in the frequency domain—achieving excellent performance with PSNR≈41 and SSIM≈0.9 [3]. Furthermore, in few-shot or even no-reference scenarios, the DeCompress algorithm en-ables denoising without relying on real clean images—using only a single noisy image—and has achieved remarkable results in combating over-fitting and zero-shot supervision scenarios [8]. Another 2024 work proposed a strategy that uses noise translation—mapping complex real-world noise to Gaussian noise—followed by denoising via Gaussian pretrained models, thereby improving generalization ability for noise far from the training distribution [9].

Against this backdrop, this study proposes a novel network architecture—the Dual-Domain Interactive Fusion Network (DDIFNet). This network is designed to address restoration challenges in specific domains, rather than merely pursuing performance metrics on general datasets. Its design philosophy involves constructing parallel processing pathways for the spatial and frequency do-mains across multiple network scales, along with introducing an innovative interactive fusion mechanism. This mechanism aims to facilitate information exchange and complementarity between feature maps of the two domains, with the goal of enabling the network to better balance the preservation of periodic patterns in forensic footprints (benefiting from frequency-domain analysis) and the restoration of random wear details (aided by spatial-domain processing). The main contributions of this study are summarized as follows:

1. Proposing DDIFNet, a network architecture synergistically processing spatial and frequency domain features in a unified framework, is pro-posed to offer a new approach for tackling image restoration challenges in professional domains.
2. Designing the core Dual-Domain Fusion Module (DDFM) and its interactive fusion mechanism, which facilitates bidirectional guidance of cross-domain information and demonstrates application potential in separating and restoring mixed structured and unstructured image details.
3. Conducting extensive experiments on a self-constructed forensic footprint dataset, which show that—compared with several mainstream comparative methods—DDIFNet achieves compe-titive performance in restoring critical evidential details and exhibits strong generalization ability on general benchmarks.

II. Related work

A. Denoising Methods Based on CNN

Since DnCNN [35] successfully applied residual learning to image denoising, a series of CNN-based methods have been proposed successively. UNet [29] and its variants utilize the encoder-decoder structure and skip connections to achieve effective fusion of multi-scale features, and have been widely used in various image restoration tasks. To handle more complex real-world noise, CBDNet [16] designed a noise estimation sub-network to solve the blind de-noising problem. In addition, methods such as RIDNet [11] have improved the network's ability to focus on key feature regions by introducing attention mecha-nisms [20]. In 2023, a feature-enhanced denoising network (FEDNet) that combines CNN and Trans-former architectures was proposed, which improves denoising performance by fusing the advantages of the two types of networks [36]. Although these methods have achieved remarkable performance on general natural image datasets, their architectural designs are usually not specifically optimized for the preservation of tiny detail-ed textures in industrial or forensic applications [4].

B. Denoising Methods Based on Transformer

Drawing on its success in the field of natural language processing [32], the Transformer architecture has also been introduced into computer vision [14]. By adopting a window-based self-attention mechanism, SwinIR [24] effectively models global dependencies in image restoration tasks and achieves state-of-the-art performance, making it an influential benchmark mo-del in this field. Taking this further, SwinIA [37] applies the Swin Transformer architecture to self-supervised blind-spot image denoising, emerging as the first end-to-end self-supervised denoising model fully based on Transformer. Additionally, studies have validated the advantages of SwinIR in multi-delay 3D ASL image denoising tasks within medical imaging scenarios [38]. However, the powerful modeling ca-pabilities of such models are primarily designed for general applicability; for domains requiring highly specialized knowledge (e.g., distinguishing shoe print wear features from background noise), their general inductive biases may not be the optimal choice.

C. Methods Incorporating Frequency-Domain Information

Several studies have attempted to integrate frequency-domain information into denoising networks. For instance, Fast Fourier Convolution [12] proposes replacing part of spatial-domain convolutions with frequency-domain convolutions to obtain a global receptive field. However, existing methods usually treat frequency-domain processing as an independent preprocessing/post-processing module or a one-way auxiliary information stream. Some studies have explored attention mechanisms in the frequency domain using Complex-Valued CNNs (CV-CNNs) to enhance spectral detail preservation and super-resolution reconstruction performance [39]. Another study focusing on infrared scanning image processing achieves denoising by predicting Fourier coefficients of periodic noise in the frequency domain [40]. Additionally, for remote sensing image enhancement, the Dual-Domain Feature Fusion Net-work (DFFN) enables collaborative spatial-frequency denoising through phased learning and fusion of amplitude and phase information [41].

Notably, these methods rarely involve dynamic, bidirectional interactive fusion mechanisms at deep network layers and across multiple scales. Yet such mechanisms may be crucial for adaptively handling mixed periodic and random features—such as those in forensic footprint images. Against this backdrop, DDIFNet is proposed to explore deeper synergy between the spatial and frequency domains, aiming to address such domain-specific challenges.

III. METHODOLOGY

IDDIFNet adopts a well-established encoder-decoder architecture [29] to leverage multi-scale information. Its core innovation lies in the designed Dual-Domain Fusion Module (DDFM), which is embedded into each layer of the encoder to replace standard convolutional blocks

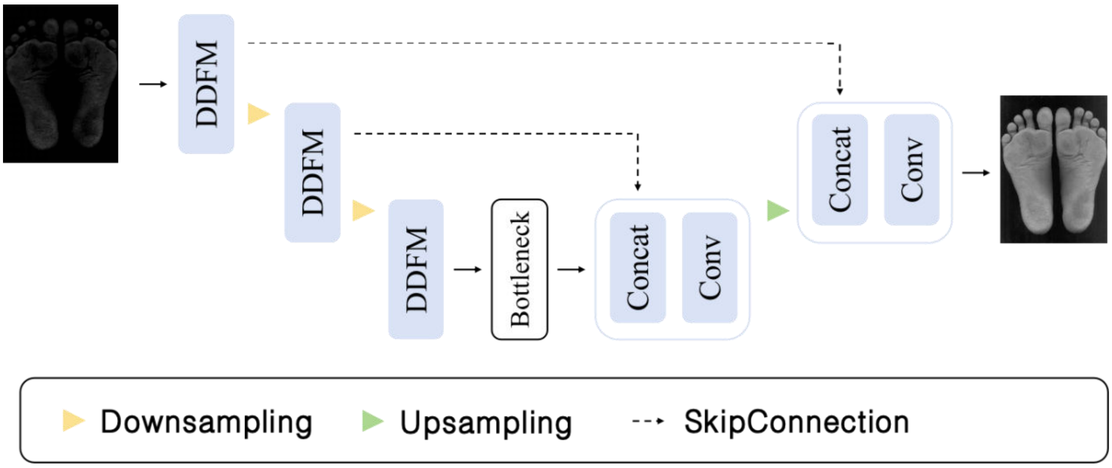

Fig.1 Overall Architecture Diagram of DDIFNet

A. Dual-Domain Fusion Module(DDFM)

As the core component of DDIFNet, the Dual-Domain Fusion Module (DDFM) aims to achieve effective extraction and interaction of spatial-domain and frequency-domain features. As illustrated in Figure 2, the input feature map Fin of the DDFM is fed into two parallel branches: the spatial-domain branch and the frequency-domain branch.

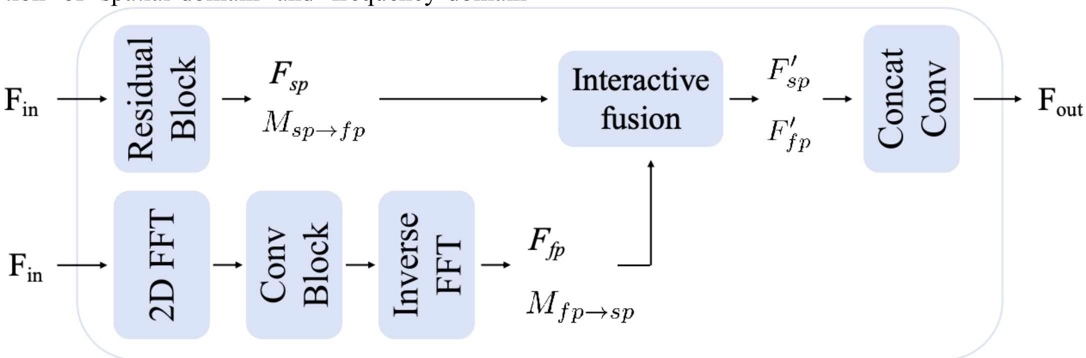

Fig.2 Detailed Structural Diagram of the Dual-Domain Fusion Module (DDFM)

Spatial-domain branch: Composed of a series of standard residual convolutional blocks [17], this branch focuses on extracting local structural and texture information. It aims to capture fine image details, edge contours, and unstructured features such as footprint wear marks. Its output is the spatial feature map Fsp.

Frequency-domain branch: This branch first transforms the input feature Fin into the frequency domain via 2D Fast Fourier Transform [13]. It then processes the real and imaginary parts of the frequency spectrum separately, utilizing a small CNN to learn how to filter and enhance features in the frequency domain. After processing, the features are transformed back to the spatial domain through inverse Fast Fourier Transform. This branch excels at identifying and handling global or periodic patterns, such as regular sole patterns and periodic noise introduced by certain sensors.

B. Interactive Fusion Mechanism

The essence of the DDFM lies in the fact that its two branches do not operate independently; instead, they exchange information through the designed interactive fusion mechanism, which is inspired by the attention concept [34]. The fusion process consists of two steps:

1. Spatial-guided frequency: The feature Fsp extracted by the spatial branch is used to generate a dynamic channel attention mask [20]. This mask is applied to the frequency-domain processed feature Ffp to reweight its channels. This enables the network to learn strategies similar to: "If the current region is a random wear area (determined by Fsp, then reduce the suppression of high frequencies in the frequency domain to preserve details."

$$F'_{fp} = M_{sp\rightarrow fp}(F_{sp}) \odot F_{fp}$$

2. Frequency-guided spatial: Similarly, we utilize the frequency-domain processed feature Ffp to generate a spatial attention map, which acts on the feature Fsp of the spatial branch. This enables the network to achieve a function similar to: "If clear periodicity of sole patterns is detected (determined by Ffp), then enhance the sharpening of corresponding structural edges in the spatial domain."

$$F'_{sp} = M_{fp\rightarrow sp}(F_{fp}) \odot F_{sp}$$

Finally, the two bidirectionally enhanced feature maps Fsp and Ffp are concatenated and fused via a 1×1 convolution, yielding the final output Fout of the DDFM module.

C. Loss Function

The Charbonnier loss function is employed to optimize the network. As a smooth approximation of the L1 loss, it helps generate high-quality images with fewer visual artifacts [22].

$$L = \sqrt{||I_{restored} - I_{clean}||^2 + \epsilon^2}$$

Here, ϵ denotes a very small constant.

IV. Experiments

A. Experimental Setup

Primary Dataset: Forensic Footprint Denoising Dataset (FFDD): To verify the model's capability in address-ing the specific problem, we constructed a Footprint Denoising Dataset (FDD). Based on the dataset used in the study by Khokher et al. [21], we selected high-quality footprint images with clear textures as clean references. Subsequently, we added complex noises that simulate real-world conditions to these images, including a mixture of Poisson noise (signal-depen-dent) and Gaussian noise. This mixture is designed to mimic low-light conditions and inherent electronic noi-se from sensors, making the dataset highly challenging and imposing strict requirements on the model's detail preservation ability.

Dataset for Generalization Test: To evaluate the model's generality, we also conducted tests on the validation set of the widely used Smartphone Image Denoising Dataset (SIDD) [10].

Evaluation Metrics: We adopted Peak Signal-to-Noise Ratio (PSNR) [19] and Structural Similarity Index Measure (SSIM) [33] as the main quantitative evalu-ation metrics. For the FDD dataset, we placed greater emphasis on SSIM, as it better reflects the recovery quality of structures and textures.

Implementation Details: The model was implemented based on the PyTorch framework [28]. The AdamW optimizer [26] was used for training, and a cosine annealing strategy [25] was employed to adjust the learning rate.

B. Performance on the Self-Constructed Dataset

Table.1 Quantitative Comparison Results on the FDD Test Set

Methods	PSNR (dB) ↑	SSIM ↑
UNet[29]	28.15	0.832
RIDNet[11]	29.33	0.881
CBDNet[16]	29.51	0.889
SwinIR[24]	30.12	0.903
DDIFNet (this work)	30.86	0.925

As can be seen from the quantitative results in Table 1, DDIFNet outperforms all comparative methods comprehensively on the challenging FFDD dataset. Notably, DDIFNet achieves a particularly significant lead in the SSIM metric—surpassing the second-best method (SwinIR) by more than 0.02. This strongly demonstrates the unique advantage of our dual-domain interaction design in preserving critical structural and texture information. Although general-purpose models like SwinIR exhibit strong performance, they still suffer from noticeable detail loss in their restored results when dealing with such task-specific textures.

Qualitative Analysis: As shown in Figure 3, the qualitative comparison results intuitively demonstrate the overwhelming superiority of DDIFNet in restoration capability. The rich skin textures and fine wrinkles in the original high-definition footprints (GT) almost completely disappear after the addition of severe noise (noisy images), rendering the images devoid of any forensic value. Baseline methods such as UNet, RIDNet, and CBDNet can remove part of the noise, but at the cost of severe image blurring and detail loss—their restored results are overly smooth and lack usable texture information. As a powerful general-purpose model, SwinIR successfully eliminates most noise and restores the main contour of the foot. However, its key limitation lies in over-smoothing: in pursuit of image purity, it erases a large number of shallow skin textures and tiny wear marks that are crucial for forensic comparison, giving the sole an unnatural "plastic-like appearance". In sharp contrast, the restoration effect of DDIFNet is visually surprisingly close to the original high-definition image. It not only removes noise perfectly but also reconstructs the finest texture network of the sole, the sharp edges of every wrinkle, and the unique wear features in a convincing manner.

Fig.3 Comparison of Restoration Effects on the Self-Constructed Footprint Dataset

C. Generalization Ability Analysis

While DDIFNet is designed to address a specific challenge, evaluating its performance in general scenarios is equally important. We conducted tests on the validation set of SIDD (Smartphone Image De-noising Dataset).

Methods	PSNR (dB) ↑	SSIM ↑
RIDNet[11]	39.25	0.954
CBDNet[16]	39.36	0.955
SwinIR[24]	39.51	0.957
DDIFNet (this work)	39.42	0.956

Table.2 Comparison of Generalization Performance on the SIDD Validation Set

As shown in Table 2, DDIFNet also achieves highly competitive performance on the general SIDD benchmark. Its results are slightly lower than those of SwinIR (which is optimized for general datasets) but outperform strong baselines such as RIDNet and CBDNet. This is a remarkably reasonable and positive outcome: it indicates that while our architecture prioritizes performance in the specific domain (forensic footprint denoising), it does not come at the cost of sacrificing generalization ability. This balance between "specialization" and "generalization" precisely demonstrates the robustness and effectiveness of our dual-domain fusion design.

D. Ablation Study

To verify the effectiveness of the DDFM (Dual-Domain Fusion Module) and its interactive fusion mechanism, we conducted ablation experiments on the FFDD dataset—this setting best aligns with the original intention of our design.

For verifying the effectiveness of the frequency-domain branch, we removed all frequency-domain branches from DDIFNet, degrading it into a pure spatial-domain CNN. This operation helps confirm whether the introduction of the frequency-domain branch contributes to the model's ability to preserve structural and texture information in forensic footprint images.

To evaluate the effectiveness of the interactive fusion mechanism, we retained both the spatial-domain and frequency-domain branches of DDIFNet but elimi-nated the interactive fusion mechanism between them. Instead, we only performed simple concatenation of the two branches at the final stage, followed by fusion via a 1×1 convolution. This scheme is intended to test whether the bidirectional interactive guidance between domains is necessary for improving the model's denoising and detail recovery performance.

Table.3 Ablation Study Results of DDIFNet on the FDD Dataset

Model Configurations	PSNR (dB)	SSIM
Spatial-Domain-Only Branch	29.28	0.879
Dual-Domain with No Interactive Fusion	30.05	0.901
Full DDIFNet	30.86	0.925

The results of the ablation experiments demonstrate that removing the frequency-domain branch leads to a significant performance degradation, particularly in terms of the SSIM metric. This indicates that frequency-domain analysis is crucial for capturing the structured patterns of footprints. Similarly, eliminating the interactive fusion mechanism also results in a noticeable performance loss, which strongly confirms that the bidirectional guidance mechanism we designed is the key to performance improvement—it enables the synergistic enhancement of the two domains rather than a mere simple feature superposition.

V. DISCUSSION

A. Core Findings

This study proposes and validates a network architecture—DDIFNet—designed to address specific high-difficulty image restoration tasks. Experimental results demonstrate that through parallel processing and deep interactive fusion in the spatial and frequency domains, DDIFNet exhibits effectiveness in restoring images severely contaminated by noise and containing complex textures. In particular, in its target application (forensic footprint analysis), its performance has achieved significant improvements compared to several general-purpose SOTA models

B. *Interpretation and Analysis of Results*

The excellent performance of DDIFNet on the FFDD dataset is attributable to the high alignment between its architectural design and the characteristics of the task. The difficulty in forensic footprint image restoration lies in distinguishing between noise and two signals with distinctly different properties: periodic sole patterns and unstructured wear marks. The frequency-domain branch of the network is inherently suitable for capturing and enhancing periodic signals, while the spatial-domain branch can finely characterize local unstructured details. More critically, the designed interactive fusion mechanism functions as a dynamic information regulator—it enables the network to adaptively balance and fuse dual-domain information based on local image content, thereby achieving the collaborative preservation of the two key forensic features. In contrast, although general-purpose models such as SwinIR exhibit strong performance, their unified self-attention mechanism may face challenges in making such fine-grained distinctions, or lead to improper handling of certain subtle yet critical details.

C. *Comparison with Existing Studies*

Compared with pure CNN methods like RIDNet, DDIFNet overcomes the locality limitation of convolutional operations to a certain extent by introducing a global perspective in the frequency domain. In contrast to SwinIR, DDIFNet's design incorporates stronger task priors and decouples the task via a dual-domain parallel approach, endowing it with greater competitiveness in the specific task of forensic footprint restoration. Notably, DDIFNet's performance on the general benchmark SIDD is slightly lower than that of SwinIR. This result is not a design flaw but rather highlights its design philosophy: balancing the pursuit of extreme professional performance and the maintenance of sound generalization ability.

D. *Limitations and Future Outlook*

Despite the promising performance exhibited by DDIFNet, the FFT/iFFT operations it incorporates and its dual-branch design undoubtedly increase the model's computational complexity. Future research may explore more lightweight dual-domain fusion modules or leverage techniques such as model distillation [18] to reduce its deployment costs. Additionally, extending this design paradigm to other restoration tasks with unique physical characteristics—such as removing specific artifacts in microscopic images—will be a valuable research direction.

VI. CONCLUSION

This study proposes a dual-domain interactive fusion network (DDIFNet) for real-world image restoration, whose design aims to address high-difficulty restoration challenges in specific professional fields. By virtue of a core module that enables parallel feature processing in the spatial and frequency domains and realizes bidirectional information guidance, the network can effectively separate and restore complex image details mixed with noise. On a self-constructed and highly challenging forensic footprint dataset, DDIFNet achieves significant performance advantages over several mainstream methods including SwinIR, while also demonstrating competitive generalization ability on general benchmarks. This work indicates that the multi-domain deep interactive fusion network designed for specific problems is an effective approach to promoting image restoration technology in solving bottlenecks in key fields and advancing toward broader practical applications.

REFERENCES

[1] Jiang, B., Li, J., Lu, Y., Cai, Q., Song, H., Lu, G. Eficient image denoising using deep learning: A brief survey. Information Fusion, 2025, 103013. https://doi.org/10.1016/j.inffus.2025.103013

[2] Mao, J., Sun, L., Chen, J., Yu, S. Overview of Research on Digital Image Denoising Methods. Sensors, 2025, 25(8), 2615.

[3] Li, B., Zhang, Y., Chen, W., Ma, Y., Li, L. Neural Network-Based Investigation of Periodic Noise Reduction Methods for High-Resolution Infrared Line Scanning Images. Remote Sensing, 2024, 16(5), 841.

[4] Zhuge, R., Wang, J., Xu, Z., Xu, Y. Single image denoising with a feature-enhanced network. Neural Networks, 2023, 168, 313-325. https://doi.org/10.1016/j.neunet.2023.08.056

[5] Fan, H., et al. Wavelet-domain frequency-mixing transformer unfolding network for low-dose computed tomography image denoising. Quantitative Imaging in Medicine and Surgery, 15(8), 7419.

[6] Wang, T., Hu, Z., Guan, Y. An efficient lightweight network for image denoising using progressive residual and convolutional attention feature fusion. 2024, Sci Rep 14, 9554. https://doi.org/10.1038/s41598-024-60139-x

[7] Zhao, X., Zhao, C., Hu, X., Zhang, H., Tai, Y., Yang, J. Learning Multi-scale Spatial-frequency Features for Image Denoising. 2025, arXiv preprint arXiv:2506.16307.

[8] Zafari, A., Chen, X., Jalali, S. DeCompress: Denoising via Neural Compression. arXiv preprint arXiv, 2025, 2503, 22015.

[9] Ha, I., Ryou, D., Seo, S., Han, B. Learning to Translate Noise for Robust Image Denoising. arXiv preprint arXiv, 2024, 2412.04727.

[10] Abdelhamed, A., Lin, S., Brown, M. S. A high-quality denoising dataset for smartphone cameras. In Proceedings of the IEEE conference on computer vision and pattern recognition , 2018, 1692-1700.

[11] Anwar, S., Barnes, N. Real image denoising with feature attention. In Proceedings of the IEEE/CVF international conference on computer vision , 2019, 3155-3164.

[12] Chi, L., Jiang, B., Mu, Y. Fast fourier convolution. Advances in Neural Information Processing Systems, 2020, 33, 4479-4488.

[13] Cooley, J. W., Tukey, J. W. An algorithm for the machine calculation of complex Fourier series. Mathematics of computation, 1965, 19(90), 297-301.

[14] Dosovitskiy, A., et al. An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv: 2020, 2010.11929.

[15] Jebur, R. S., Zabil, M. H. B. M., Hammood, D. A., Cheng, L. K. A comprehensive review of image denoising in deep

learning. Multimedia Tools and Applications, 2024, 83(20), 58181-58199.
[16] Guo, S., Yan, Z., Zhang, K., Zuo, W., Zhang, L. Toward convolutional blind denoising of real photographs. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2019, 1712-1722.
[17] He, K., Zhang, X., Ren, S., Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, 770-778.
[18] Hinton, G., Vinyals, O., Dean, J. Distilling the knowledge in a neural network. arXiv preprint arXiv, 2015, 1503.02531.
[19] Hore, A., Ziou, D. Image quality metrics: PSNR vs. SSIM. In 2010 20th international conference on pattern recognition, IEEE, 2010, 2366-2369.
[20] Hu, J., Shen, L., Sun, G. Squeeze-and-excitation networks. In Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, 7132-7141.
[21] Khokher, R., Singh, R. C., Kumar, R. Footprint recognition with principal component analysis and independent component analysis. In Macromolecular symposia, 2015, Vol. 347, No. 1, 16-26.
[22] Lai, W. S., Huang, J. B., Ahuja, N., Yang, M. H. Deep laplacian pyramid networks for fast and accurate super-resolution. In Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, 624-632.
[23] LeCun, Y., Bengio, Y., Hinton, G. Deep learning. nature, 2015, 521(7553), 436-444.
[24] Liang, J., Cao, J., Sun, G., Zhang, K., Van Gool, L., Timofte, R. Swinir: Image restoration using swin transformer. In Proceedings of the IEEE/CVF international conference on computer vision, 2021, 1833-1844.
[25] Loshchilov, I.,Hutter, F. Sgdr: Stochastic gradient descent with warm restarts. arXiv preprint arXiv, 2016, 1608.03983.
[26] Loshchilov, I., Hutter, F. Decoupled weight decay regularization. arXiv preprint arXiv, 2017, 1711.05101.
[27] Oppenheim, A. V. Discrete-time signal processing. Pearson Education India. 1999.
[28] Paszke, A. , Gross, S. , Massa, F. , Lerer, A. , Chintala, S. Pytorch: an imperative style, high-performance deep learning library. 2019.
[29] Ronneberger, O., Fischer, P., Brox, T. U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical image computing and computer-assisted intervention. Cham: Springer international publishing. 2015, 234-241.
[30] Ross, L. The Image Processing Handbook, Russ John C. CRC Press, Boca Raton FL, 2011, 972 pages. ISBN 1-4398-4045-0 (Hardcover).
[31] Tian, C., Xu, Y., Zuo, W. Image denoising using deep CNN with batch renormalization. Neural Networks, 2020, 121, 461-473. https://doi.org/10.1016/j.neunet.2019.08.022
[32] Vaswani, A., et al. Attention is all you need. Advances in neural information processing systems, 2017, 30.
[33] Wang, Z., Bovik, A. C., Sheikh, H. R., Simoncelli, E. P. Image quality assessment: from error visibility to structural similarity. IEEE transactions on image processing, 2004, 13(4), 600-612.
[34] Woo, S., Park, J., Lee, J. Y., Kweon, I. S. Cbam: Convolutional block attention module. In Proceedings of the European conference on computer vision (ECCV), 2018, 3-19.
[35] Zhang, K., Zuo, W., Chen, Y., Meng, D., Zhang, L. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. IEEE transactions on image processing, 2017,26(7), 3142-3155.
[36] Zhuge, R., Wang, J., Xu, Z., Xu, Y. Single image denoising with a feature-enhanced network. Neural Networks, 2023, 168, 313-325. https://doi.org/10.1016/j.neunet.2023.08.056
[37] Papkov, M., Chizhov, P., Parts, L. Swinia: Self-supervised blind-spot image denoising without convolutions. In 2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), 2025, 7071-7080.
[38] Shou, Q.,et al. Transformer-based deep learning denoising of single and multi-delay 3D arterial spin labeling. Magnetic resonance in medicine, 2024, 91(2), 803-818.
[39] Shah, Z. H., Müller, M., Hübner, W., Ortkrass, H., Hammer, B., Huser, T., Schenck, W. Image restoration in frequency space using complex-valued CNNs. Frontiers in Artificial Intelligence, 2024, 7, 1353873
[40] Li, B., Zhang, Y., Chen, W., Ma, Y., Li, L. Neural Network-Based Investigation of Periodic Noise Reduction Methods for High-Resolution Infrared Line Scanning Images. Remote Sensing, 2024, 16(5), 841.
[41] Yao, Z., Fan, G., Fan, J., Gan, M., Chen, C. P. Spatial-frequency dual-domain feature fusion network for low-light remote sensing image enhancement. IEEE Transactions on Geoscience and Remote Sensing. 2024.