

# Spatial-Frequency Guided Moiré Removal with Multi-Stage Feature Fusion

Chen Lo 1 and Chia-Hung Yeh 2\*

National Taiwan Normal University, Taiwan

E-mail: clo20@ntnu.edu.tw

\*National Taiwan Normal University, Taiwan

Sun Yat-sen University, Taiwan

E-mail: chieh@ntnu.edu.tw

**Abstract**— Moiré patterns severely degrade the visual quality of images by introducing complex, multi-scale, and frequency-entangled artifacts that are often intertwined with real image content. To address this challenge, we propose a novel multi-stage progressive restoration framework, named SFMFNet, which jointly exploits spatial- and frequency-domain representations to enhance moiré removal. The network adopts an encoder-decoder structure with multi-scale token mixing blocks and a dedicated frequency-aware module based on Fourier transforms. To improve information flow across stages, we introduce a cross-scale feature aggregation (CSFA) mechanism that effectively aggregates and redistributes features at different resolutions. Experimental results demonstrate that SFMFNet can effectively suppress moiré artifacts while preserving structural integrity and texture details, confirming the advantage of combining spatial structure and frequency cues for generative image restoration.

## I. INTRODUCTION

Moiré patterns are complex visual distortions caused by the interference between overlapping periodic structures, such as display grids and camera sensor arrays. These artifacts often appear as colored stripes, ripples, or curved textures that span a wide range of spatial scales and frequencies, significantly degrading image quality. Unlike conventional degradations such as blur or noise, moiré patterns are frequency-entangled and highly structured, making them particularly difficult to remove. With the growing prevalence of high-resolution cameras and displays, moiré artifacts have become increasingly common, prompting the need for specialized restoration techniques. Traditional signal processing methods, such as notch filters [1] or low-rank decompositions [2], often fail to generalize across diverse moiré patterns and typically lead to oversmoothing or incomplete removal. In contrast, deep learning-based approaches have shown remarkable success in capturing the complex statistical and structural characteristics of moiré interference. Motivated by the frequency-domain nature of moiré artifacts, we analyze the Fourier transforms of moiré-contaminated and clean images and observe that the degradation mainly resides in the amplitude spectrum, while

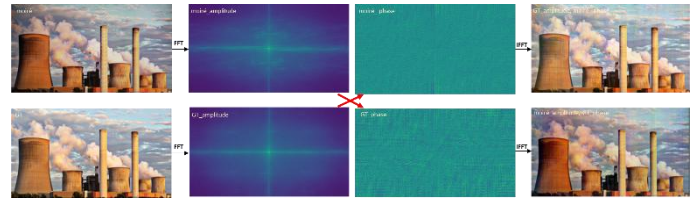


Fig. 1 Comparison between two frequency-domain recombination: (top) GT amplitude + moiré phase and (bottom) moiré amplitude + GT phase. The results reveal that moiré artifacts are primarily encoded in the amplitude spectrum, underscoring the importance of amplitude-aware learning for moiré removal.

the phase spectrum largely retains structural information. As illustrated in Fig. 1, replacing the moiré phase with a clean one has limited effect, whereas substituting clean amplitude effectively suppresses moiré, even when the phase remains contaminated, highlighting the dominant role of amplitude in moiré formation. This observation suggests that frequency-aware learning strategies, particularly those focusing on amplitude modulation, may offer a more effective and interpretable solution for moiré removal. To this end, we propose a multi-stage progressive restoration framework that suppresses moiré patterns through spatial-frequency fusion. Our model integrates FFT-guided modules and multiscale refinement blocks to selectively attenuate undesired frequency components while preserving fine-grained high-frequency textures, ultimately achieving faithful restoration with improved visual fidelity.

## II. RELATED WORK

An overview of recent developments in moiré pattern removal is presented in this section, with a specific focus on deep learning-based techniques. The discussion begins by examining various deep neural network architectures designed for moiré removal, highlighting their core strategies such as multiscale processing, frequency-aware attention, and progressive restoration. These approaches have laid the foundation for significant improvements in visual quality and artifact suppression. The second part of this section explores the integration of Fourier transform into deep learning frameworks, emphasizing how frequency-domain

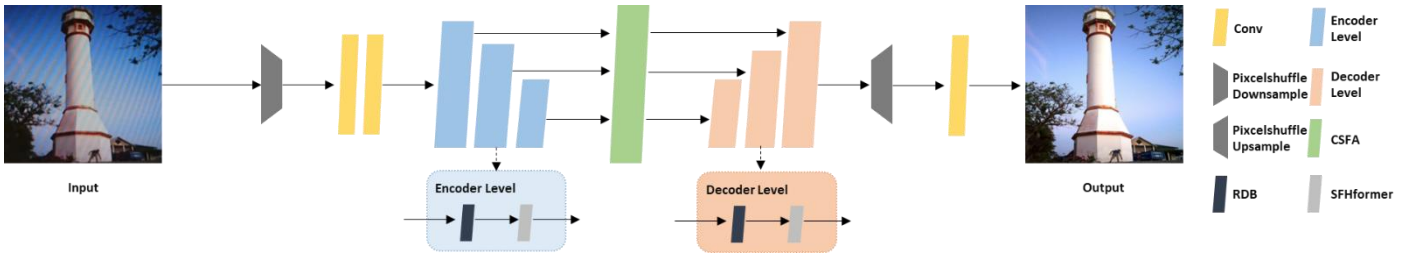


Fig. 2 Architecture of the proposed spatial-frequency guided moiré removal network (SFMNet).

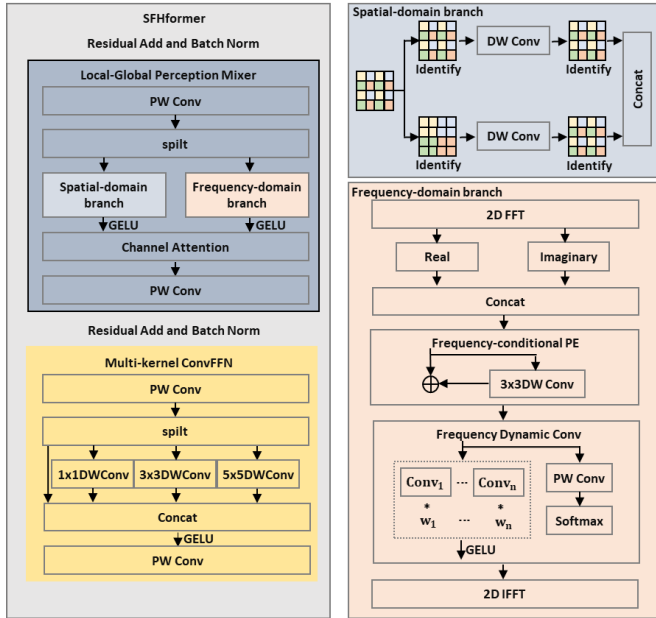


Fig. 3 Architecture of the SFHformer block [14].

representations can enhance the network’s ability to identify and suppress moiré artifacts. Together, these two perspectives provide critical insights that inform the design and optimization of our proposed method.

#### A. Moiré Removal by Deep Learning

Recent progress in deep learning has significantly advanced the handling of various moiré artifacts. Sun et al. [3] introduced DMCNN, a multiscale convolutional network designed to suppress moiré patterns across diverse frequency ranges. Alongside this, they released the TIP dataset, a large-scale benchmark curated from ImageNet to support training and evaluation. Building upon a coarse-to-fine paradigm, Liu et al. [4] proposed a two-branch CNN architecture: the coarse branch performs downsampling and initial moiré suppression via residual learning, while the fine branch focuses on upsampling and output refinement. Addressing both frequency misalignment and chromatic distortions, He et al. [5] developed MopNet, which leverages multiscale feature aggregation, color channel analysis based on edges, and a perceptual classification module to model the visual diversity of moiré. To handle high-resolution scenarios, Yu et al. [6] presented UHDM, a detection framework tailored for 4K UHD images.

Further improvements were made by Yang et al. [7], who proposed a parallel multiscale CNN enhanced with residual

blocks, an Information Exchange Module (IEM), and a Final Feature Fusion (FFF) block. Their design also includes a scale-aware module aligned with semantic features to better resolve multi-resolution moiré patterns. In a subsequent work, Yang et al. [8] introduced a moiré pattern modeling technique that enables precise alignment of synthetic training triplets, significantly benefiting their MoireDet detection network.

Niu et al. [9] adopted a progressive restoration strategy, where moiré artifacts are removed hierarchically from low to high resolution in a multi-stream network. They also proposed a texture complementation module that merges intermediate features across streams to enhance fine-detail recovery. Lastly, Nguyen et al. [10] focused on moiré removal in screenshot images, proposing a multiscale guided demoiréing network. Their method utilizes multiscale feature extractors, frequency transformation modules (FTBs), and guided residual blocks (MGRBs) to progressively refine moiré-related representations across the frequency spectrum.

#### B. Fourier Transform in Neural Networks

The Fourier transform has been increasingly adopted in neural networks due to its ability to capture global context and decompose features across different frequency components. Early works explored frequency-domain representations in high-level vision tasks to model non-local relationships, such as fast fourier convolution (FFC) [11] and global filter networks [12]. With the advancement of frequency-based analysis, recent studies have demonstrated that frequency features are also highly effective for low-level vision tasks like image inpainting, deraining, and super-resolution.

Chen et al. [13] proposed the Focal Frequency Loss to guide models to focus on frequency components with larger reconstruction errors, leading to more precise restoration of structural and textural details in image generation and enhancement tasks. SFHformer [14] incorporates FFT into a Transformer-style framework, combining spatial and frequency branches with frequency-aware modules to enable efficient global modeling and strong restoration performance across diverse degradation types. These studies highlight the growing potential of frequency-domain modeling in improving restoration quality while maintaining computational efficiency

### III. PROPOSED METHOD

In this section, we present the architecture of the proposed moiré removal framework, named SFMFNet, designed to effectively suppress frequency-rich moiré artifacts and recover high-fidelity images. The architecture consists of a hierarchical encoder, a frequency-aware token mixing module, cross-scale feature aggregation (CSFA) modules, and a progressive decoder. The overall structure is illustrated in Fig. 2.

#### A. Overall Architecture

SFMFNet takes an RGB image as input and outputs a restored image of the same resolution. The framework follows a symmetric encoder-decoder design with three resolution stages that operate progressively from coarse to fine. The encoder is responsible for extracting multi-scale spatial-frequency features through residual dense blocks and SFHformer. These features are then enhanced by CSFA modules, which facilitate effective cross-scale communication and preserve both global and local information. The refined features are passed into the decoder, where they are gradually upsampled and reconstructed via skip connections and convolutional refinement blocks. Finally, the output feature is projected back to the image space, and a residual connection adds the result to the input image to preserve the overall image structure.

#### B. Hierarchical Encoder with SFHformer

The encoder is designed to extract multi-scale representations by combining residual dense blocks (RDBs) and SFHformer blocks[14], as illustrated in Fig. 3. Inspired by the original SFHformer architecture, each block integrates local spatial convolution with global frequency modeling through Fourier transforms. This design enables the network to effectively capture both fine-grained textures and frequency-specific moiré distortions. To maintain the lightweight and modular nature of our framework, we adopt and slightly simplify the original token mixing mechanism while preserving its core functionality.

Given an input image  $x \in \mathbb{R}^{3 \times H \times W}$ , the encoder generates three feature maps:

$$\begin{aligned} y_1 &= SFHformer_1(RDB_1(x)), \\ y_2 &= SFHformer_2(RDB_2(\mathcal{D}(y_1))), \\ y_3 &= SFHformer_3(RDB_3(\mathcal{D}(y_2))), \end{aligned} \quad (1)$$

where  $RDB$  represents a residual dense block,  $SFHformer$  refers to our spatial-frequency unit, and  $\mathcal{D}$  denotes a downsampling operation.

#### C. Cross-scale Feature Aggregation

To improve feature interaction across stages, we insert a dedicated CSFA module after each encoding level. Each CSFA module receives a set of multi-scale features and performs

feature aggregation and redistribution via residual convolution, attention-based feature recalibration, and up/downsampling operations. The fused features are denoted as:

$$\begin{aligned} o_1 &= CSFA_1(y_1, \mathcal{U}(y_2), \mathcal{U}^2(y_3)), \\ o_2 &= CSFA_2(\mathcal{D}(y_1), y_2, \mathcal{U}(y_3)), \\ o_3 &= CSFA_3(\mathcal{D}^2(y_1), \mathcal{D}(y_2), y_3), \end{aligned} \quad (2)$$

where  $\mathcal{U}$  denotes an upsampling operation, and  $\mathcal{D}$  denotes a downsampling operation. This design ensures that each CSFA module integrates information from both coarser and finer levels, allowing the network to enhance consistency and preserve structural cues across scales. These refined features  $o_1, o_2, o_3$  are passed to the decoder for final reconstruction.

#### D. Progressive Decoder with Residual Reconstruction

The decoder reconstructs the output in a coarse-to-fine manner. At each stage, it combines CSFA-refined features with skip-connected encoder outputs. Upsampling is achieved via pixel shuffle, followed by residual refinement. The reconstruction process can be expressed as:

$$\begin{aligned} d_3 &= \mathcal{U}(SFUnit_3(RDB_3(o_3))), \\ d_2 &= \mathcal{U}(SFUnit_2(RDB_2([d_3, o_2]))), \\ d_1 &= \mathcal{U}(SFUnit_1(RDB_1([d_2, o_1]))), \end{aligned} \quad (3)$$

where  $d_i$  is the reconstructed feature at stage  $i$ , and  $[\cdot, \cdot]$  denotes channel-wise concatenation. Finally, the output feature is projected back to the image space, and a residual connection adds the result to the input image:

$$\hat{x} = Conv(d_1) + x, \quad (4)$$

this residual design helps preserve low frequency content and accelerates convergence.

#### E. Loss Function

To train the proposed network, we formulate the task as a mapping from a moiré-contaminated image  $x$  to its corresponding clean target  $x_{gt}$ . The network output  $\hat{x}$  is supervised using a combination of pixel-wise  $\mathcal{L}_1$  loss, perceptual loss, and frequency-domain loss. The  $\mathcal{L}_1$  loss encourages accurate restoration at the pixel level, while the perceptual loss, computed using high-level features extracted from a pre-trained VGG network, promotes semantic consistency between the restored and ground-truth images. Additionally, a Fourier loss is applied in the frequency domain to penalize amplitude differences and suppress moiré-specific artifacts. The total loss is the sum of these components:

$$\begin{aligned} \mathcal{L}_{total} &= \mathcal{L}_1(\hat{x}, x_{gt}) + \mathcal{L}_p(\hat{x}, x_{gt}) \\ &\quad + \mathcal{L}_{fft}(\hat{x}, x_{gt}). \end{aligned} \quad (5)$$

This integrated objective guides the network to produce reconstructions with both spatial fidelity and frequency consistency.

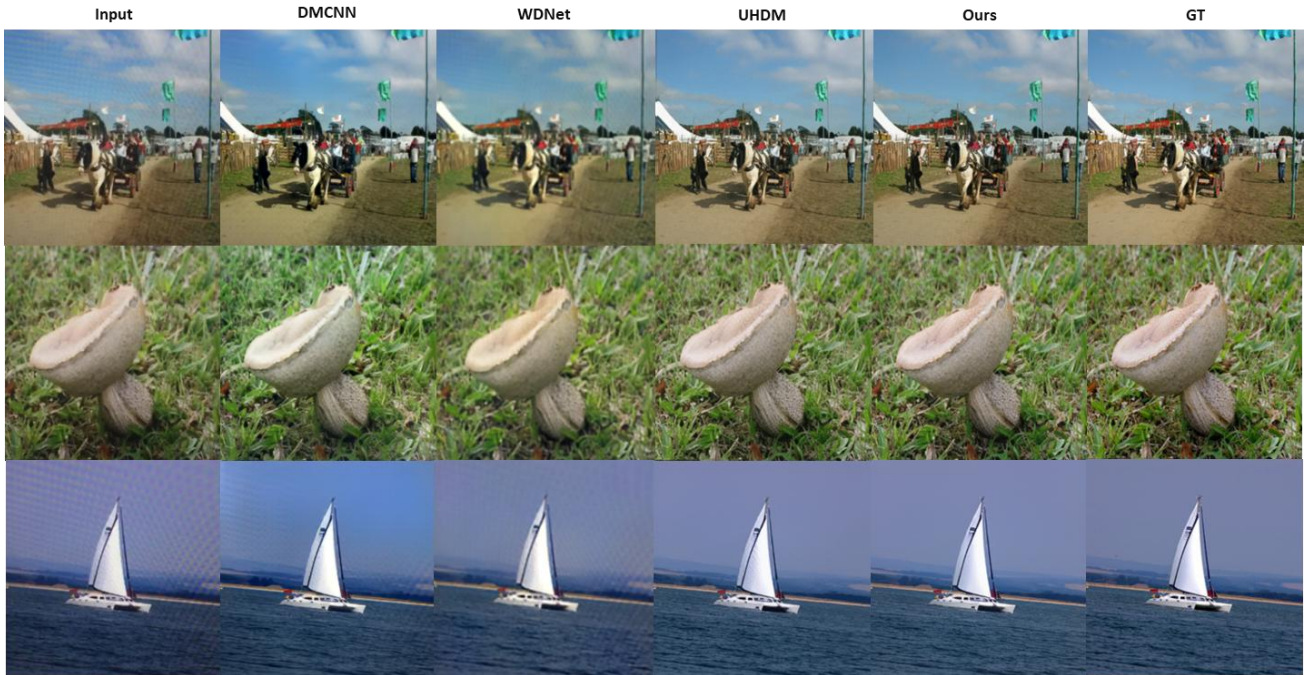


Fig. 4 Visual Comparison of DMCNN, WNet, UHDM, and the proposed method on the TIP dataset.

TABLE I

COMPARISONS OF QUANTITATIVE RESULTS EVALUATED IN TERMS OF AVERAGE PSNR AND SSIM								
Dataset	Metrics	Input	DMCNN [3]	WNet [15]	MBCNN [16]	EAFN [17]	UHDM [6]	Ours
TIP dataset	PSNR $\uparrow$	20.30	26.77	28.08	30.03	29.70	30.11	<b>30.32</b>
	SSIM $\uparrow$	0.738	0.871	0.904	0.893	0.893	0.920	<b>0.942</b>

#### IV. EXPERIMENTS RESULTS

To comprehensively evaluate the performance of our proposed SFMFNet architecture for moiré pattern removal, we conduct both quantitative and qualitative experiments. Our model is benchmarked against several state-of-the-art demoiréing approaches, including DMCNN [3], WNet [15], MBCNN [16], EAFM [17], and UHDM [6] using standard evaluation metrics

##### A. Experimental Details and Datasets

The implementation of our method is based on Python with PyTorch 2.1.1, and all experiments are conducted on a workstation equipped with an Intel i7-14700F CPU (2.9 GHz), 64 GB RAM, and an NVIDIA RTX 4090 GPU. During training, we set the input image resolution to  $256 \times 256$  and adopt a batch size of 2. The model is optimized using the Adam algorithm, starting with a learning rate of 0.0002, which is gradually decayed through cyclic cosine annealing. To evaluate our framework, we use the TIP dataset [3], consisting of 135,000 paired moiré-contaminated and clean images. Specifically, 90% of the dataset is used for training, while the remaining 10% is reserved for testing.

##### B. Quantitative Evaluation

To quantitatively evaluate the moiré removal performance, we compare our proposed model with five state-of-the-art deep

learning-based methods: DMCNN, WNet, MBCNN, EAFM, and UHDM. The evaluation is based on two widely adopted image quality metrics: Peak Signal-to-Noise Ratio (PSNR) [18] and Structural Similarity Index Measure (SSIM) [19]. As shown in Table I, our method consistently achieves better scores than the compared approaches, demonstrating its effectiveness in removing moiré artifacts while preserving fine details and image structures.

##### C. Qualitative Evaluation

Fig. 4 presents a qualitative comparison on the TIP dataset between our proposed method and several state-of-the-art demoiréing approaches, including DMCNN, WNet, and UHDM. The input images are heavily corrupted by moiré artifacts that obscure texture and structure details. While DMCNN and WNet reduce some of the moiré artifacts, they often introduce blurring or leave residual interference, especially in fine-grained areas such as grass, object contours, or repetitive textures. UHDM delivers improved results, yet some moiré traces persist, particularly in high-frequency regions. In contrast, our method demonstrates superior restoration quality as it effectively suppresses moiré patterns while preserving detailed textures, edges, and color consistency. Compared to the ground truth, our results show greater visual similarity and fewer distortions across various image scenes.

## V. CONCLUSIONS

In this paper, we proposed SFMFNet, a novel moiré pattern removal network that fuses spatial and frequency-domain information within a multi-stage architecture. The network leverages SFHformer blocks for spatial-frequency token mixing, CSFA modules for multi-scale feature interaction, and a progressive decoder for detail-preserving image reconstruction. To ensure both visual fidelity and frequency consistency, we employ a hybrid loss function combining pixel-wise, perceptual, and Fourier-based terms. Extensive experiments validate that SFMFNet achieves competitive performance, offering a robust solution for high-fidelity image restoration under complex moiré degradations.

## VI. ACKNOWLEDGMENT

This work was supported by the National Science and Technology Council, Taiwan, under Grant NSTC 113-2221-E-003-018-MY3.

## REFERENCES

- [1] Z. Wei, J. Wang, H. Nichol, S. Wiebe and D. Chapman, “A median-Gaussian filtering framework for Moiré pattern noise removal from X-ray microscopy image,” *Micron*, pp. 170–176, Feb. 2012.
- [2] J. Yang, F. Liu, H. Yue, X. Fu, C. Hou and F. Wu, “Textured Image Demoiréing via Signal Decomposition and Guided Filtering,” *IEEE Trans. Image Process.*, vol. 26, no. 7, pp. 3528–3541, July 2017.
- [3] Y. Sun, Y. Yu and W. Wang, “Moiré Photo Restoration Using Multiresolution Convolutional Neural Networks,” *IEEE Trans. Image Process.*, vol. 27, no. 8, pp. 4160–4172, Aug. 2018.
- [4] B. Liu, X. Shu and X. Wu, “Demoiréing of Camera-Captured Screen Images Using Deep Convolutional Neural Network,” arXiv:1804.03809, Apr. 2018.
- [5] B. He, C. Wang, B. Shi and L. Duan, “Mop Moiré Patterns Using MopNet,” in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, pp. 2424–2432, Oct. 2019.
- [6] X. Yu, P. Dai, W. Li, L. Ma, J. Shen, J. Li and X. Qi, “Towards efficient and scale-robust ultra-high-definition image demoiréing,” in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, pp. 646–662, Oct. 2022.
- [7] S. Yang, Y. Lei, S. Xiong and W. Wang, “High resolution demoire network,” in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, pp. 888–892, Oct. 2020.
- [8] C. Yang, Z. Yang, K. Ke, T. Chen, M. Grzegorzec and J. See, “Doing More with Moiré Pattern Detection in Digital Photos,” *IEEE Trans. Image Process.*, vol. 32, pp. 694–708, Jan. 2023.
- [9] Y. Niu, Z. Lin, W. Liu and W. Guo, “Progressive Moire Removal and Texture Complementation for Image Demoireing,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 33, pp. 3608–3621, 2023.
- [10] D. H. Nguyen, S. Lee and C. Lee, “Multiscale Coarse-to-Fine Guided Screenshot Demoiréing,” *IEEE Signal Process. Lett.*, vol. 30, pp. 898–902, 2023.
- [11] R. Suvorov, E. Logacheva, A. Mashikhin, A. Remizova, A. Ashukha, A. Silvestrov, N. Kong, H. Goka, K. Park and V. Lempitsky, “Resolution-robust large mask inpainting with fourier convolutions,” in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 2149–2159, 2022.
- [12] Y. Rao, W. Zhao, W. Zhu, J. Lu and J. Zhou, “Global filter networks for image classification,” *Advances in neural information processing systems*, 980–993, 2021.
- [13] D. Fuoli, L. Van Gool, and R. Timofte, “Fourier space losses for efficient perceptual image super-resolution,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 2360–2369, 2021.
- [14] X. Jiang, X. Zhang, N. Gao and Y. Deng, “When Fast Fourier Transform Meets Transformer for Image Restoration,” in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, pp. 381–402, 2024.
- [15] L. Liu, J. Liu, S. Yuan, G. Slabaugh, A. Leonardis, W. Zhou and Q. Tian, “Wavelet-Based Dual-Branch Network for Image Demoiréing,” in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, pp. 86–102, Nov. 2020.
- [16] B. Zheng, S. Yuan, G. Slabaugh and A. Leonardis “Image Demoireing with Learnable Bandpass Filters,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pp. 3636–3645, 2020.
- [17] C. Sun, H. Lai, L. Wang and Z. Jia, “Efficient Attention Fusion Network in Wavelet Domain for Demoireing,” *IEEE Access*, March. 2021.
- [18] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, “Image quality assessment: From error visibility to structural similarity,” *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
- [19] R. Zhang, P. Isola, A. A. Efros, E. Shechtman and O. Wang, “The Unreasonable Effectiveness of Deep Features as a Perceptual Metric,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2018.