

# Moiré Pattern Removal via Attentive Fractal Network

Dejia Xu<sup>1\*</sup> Yihao Chu<sup>2\*</sup> Qingyan Sun<sup>3</sup>

Peking University<sup>1</sup>, Beijing University of Posts and Telecommunications<sup>2</sup>, Beijing Jiaotong University<sup>3</sup>

dejia@pku.edu.cn<sup>1</sup>, ccc@bupt.edu.cn<sup>2</sup>, sunqingyanstc@bjtu.edu.cn<sup>3</sup>

## Abstract

*Moiré patterns are commonly seen artifacts when taking photos of screens and other objects with high-frequency textures. It's challenging to remove the moiré patterns considering its complex color and shape. In this work, we propose an Attentive Fractal Network to effectively solve this problem. First, we construct each Attentive Fractal Block with progressive feature fusion and channel-wise attention guidance. The network is then fractally stacked with the block on each of its levels. Second, to further boost the performance, we adopt a two-stage augmented refinement strategy. With these designs, our method wins the burst demoiréing track and achieves second place in single image demoiréing and single image deblurring tracks in NTIRE20 Challenges. Extensive experiments demonstrate the superiority of our method for moiré pattern removal compared to existing state-of-the-art methods, and prove the effectiveness of its each component. We will publicly release our code and trained weights on <https://github.com/irld/AFN>.*

## 1. Introduction

With the development of camera imaging technology, more advanced capturing devices are widely used in our daily life. However, when the image is taken in front of a screen or contains other objects with high-frequency textures, moiré patterns are likely to appear. Due to the loss in detail and distortion in signal, the visual appearance of the image is degraded. This also results in the failure of many high-level computer vision applications, since they are often built on datasets of clean images. An example of an image with moiré patterns can be observed in Fig. 1.

Moiré effect is of great use in many domains including marine navigation [6] and strain measurement [32]. But when it comes to digital photography, it causes degradation in visual quality of images. According to the Nyquist theorem [4], moiré patterns are more likely to appear when

\* indicates equal contribution.

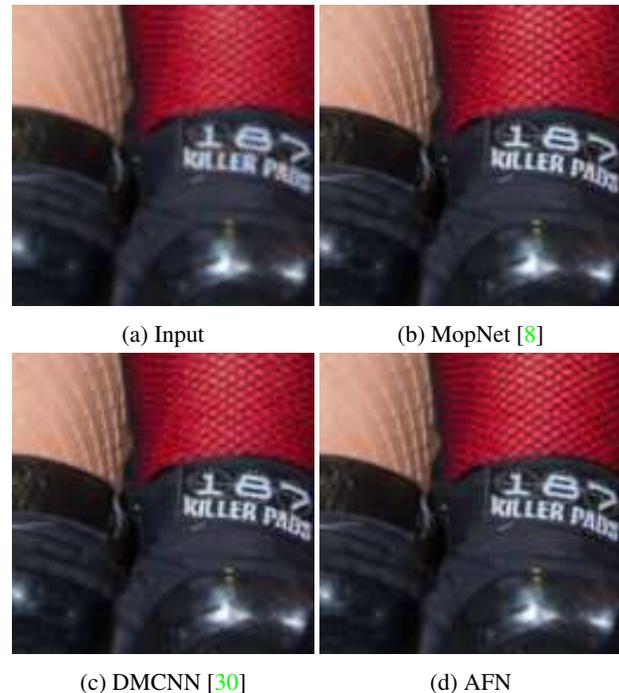


Figure 1: Visual comparison of our AFN with state-of-the-art approaches. Our AFN not only removes the artifact effectively but also preserves the details of the textures.

textures of photographed objects share a similar frequency with the color filter array. Due to the information loss, it is a highly ill-posed problem to remove the undesired moiré patterns. And considering its diversity in shape and color, such restoration gets challenging.

Camera manufacturers have adopted various hardware-level improvements [24, 33] including optical filters and specially designed color filter arrays [23, 21, 33, 31]. For the commonly-used Bayer filter [5], post-processing algorithms are more ideal solutions. Various methods have been proposed using signal processing theories [37, 2] and exploit low-rank constraint [18]. Recently, deep neural network based methods appeared and became the mainstream gradually. In [30], a multi-scale convolution network is proposed to learn the mapping from moiré pat-

tern contaminated images to clean images. He *et al.* [8] utilized edge maps and other human-labeled moiré pattern attributes to guide the removal of the moiré patterns. Though they achieved better performance than non-learning approaches, they were mainly designed for images captured from screens and are ideal when dealing with images containing other high-frequency textures.

Considering the limitations of existing works, we aim to design a deep learning framework that is capable of effectively restoring clean images from moiré pattern contamination. First, we construct an Attentive Fractal Block via progressive feature fusion and channel-wise attention guidance. Then we stack our AFB in a fractal way inspired by fractal architecture design [15]. Additionally, a two-stage augmented refinement strategy is adopted to fuse the geometric transformations of the coarse results. With the progressively fused feature representations and the power of two-stage augmented refinement, our AFN outperforms previous moiré pattern removal networks and other popular backbones. Our contributions are summarized as follows:

- We propose a fractally designed network for moiré pattern removal. With the help of progressive feature fusion and channel-wise attention guidance, the constructed Attentive Fractal Network manages to model the moiré pattern and effectively remove the artifacts.
- We propose a two-stage augmented refinement strategy to deal with the severely distorted images. The coarse results are further refined via fusing the geometric transformed outputs. Such strategy can significantly boost overall performance.
- Extensive experiments demonstrate the superiority of AFN for moiré pattern removal and deblurring task as well. A rich ablation study proves the effectiveness of each component.

## 2. Related Work

**Moiré Removal** Research on moiré pattern removal goes on for decades. Some digital camera manufacturers installed optical low pass filters on cameras [24] which reduced moiré effects but they often blurred the images in the meantime. Some digital camera manufacturers also developed new CMOS with a more moiré-resistant color filter array, such as FOVEON X3 CMOS [23, 21, 33] and Fujifilm X-Trans CMOS [31]. However, the Bayer filter [5] is still the most common color filter array used in digital cameras, to remove moiré patterns on images produced by Bayer CFA is significantly important. A few computational methods have been proposed. Wei *et al.* [37] proposed a median-Gaussian filtering framework to remove moiré patterns from X-ray microscopy image. However, it could not handle colored digital images. A Photoshop plug-in called

Sattva Descreen [2] claimed that it uses Fourier transform to precisely remove moiré patterns. But it still suffered from blurring effects. In [18], Liu *et al.* proposed a moiré removing method based on low-rank and sparse matrix decomposition in frequency domain.

With the rising computation ability and the marvelous performance of deep learning in low-level vision tasks, convolutional neural networks have been applied to removing moiré patterns. Sun *et al.* [30] proposed a multi-scale convolutional neural network to learn the mapping from moiré images to clear images. He *et al.* [8] utilized the imbalanced properties of moiré effects and presented MopNet to remove moiré patterns with the help of edge maps and other moiré pattern attributes like shape, etc. However, MopNet required extra data labels of moiré pattern attributes to achieve a higher PSNR, which is not often available in real cases. Recently, Yuan *et al.* hosted a demoiré challenge [40] and attracted more researchers on removing moiré patterns via deep learning. These deep learning networks were mainly designed and trained for removing moiré patterns from photos of screens. They don't yield a pleasant performance when it comes to moiré patterns on photos of other high-frequency textured objects such as textiles, tiles and chain meshes, etc.

**Learning Based Image Restoration** With the rapid development of Deep Learning, many techniques have been developed to tackle the problem of image restoration under various degradations. Gu *et al.* [7] proposed a Self-Guided Network for image denoising. Wei *et al.* [36] exploited Retinex Theory for low light enhancement. Qu *et al.* [28] presented a GAN framework for image dehazing. In [17], Li *et al.* proposed a rain streak removal network based on physics models. Zhang *et al.* [41] introduced a Stacked Hierarchical Network for image deblurring. In [22], Ma *et al.* adopted a Cyclic GAN to generate and separate reflections. However, less research has been made regarding the removal of moiré pattern artifacts. Existing methods for other tasks might fail to handle moiré patterns considering its diverse shape and color.

## 3. Attentive Fractal Network

**Fractal Network** Many of the hand-crafted networks are organized into a hierarchical structure. In [15], Larsson *et al.* proposed a fractal architecture, which is a very deep network without residuals. Yang *et al.* [38] introduced fractal band learning (FBL) for rain streak removal. FBL networks conduct band feature refinement, expansion and fusion operations based on band recovery theory, and capture the hierarchical band dependency with learnable components. In [14], Kwak *et al.* proposed a Fractal Residual Network and adopted RCAN [43] as its fractal design. In our work, we aim to exploit the fractal features and develop a new effective backbone for moiré pattern removal.

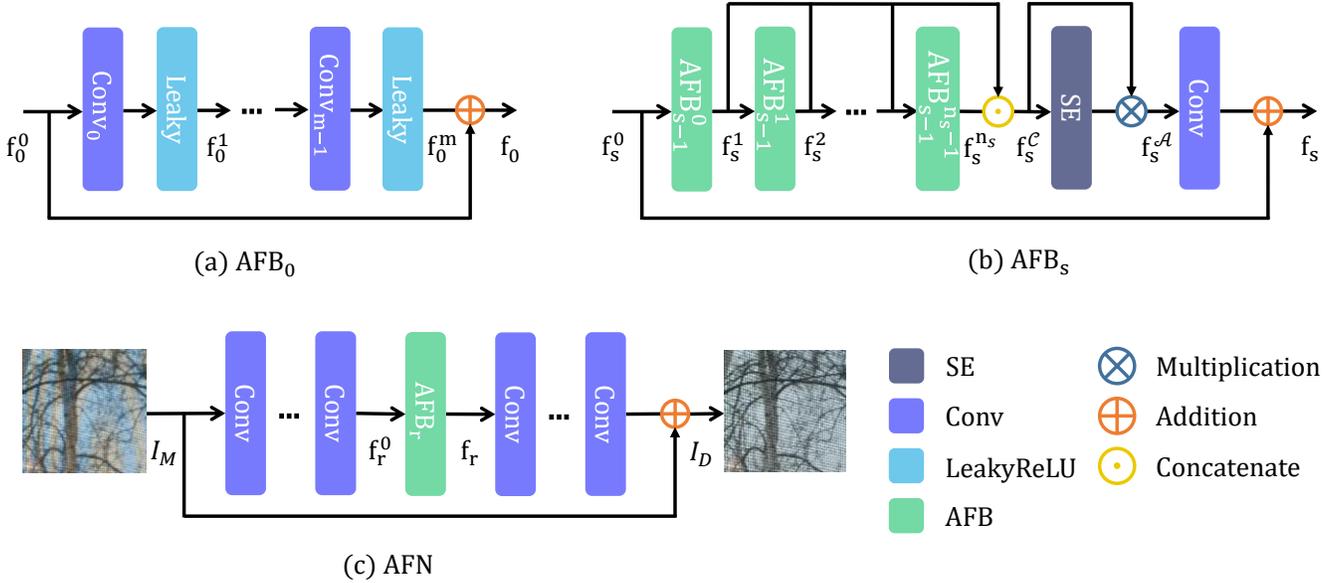


Figure 2: Architecture of our proposed framework. (a) shows the structure of the basic block  $\text{AFB}_0$ . (b) shows the hierarchical architecture of  $\text{AFB}_s$ , where  $s \in [1, r]$  is the recursion level of the block. (c) illustrates the structure of AFN, which consists of encoding layers, an  $\text{AFB}_r$ , and decoding layers from left to right.

### 3.1. Network Overview

Moiré patterns are complex in both color and shape, and restoring such a contaminated image is an ill-posed problem. A naive deep network would suffer from vanishing gradients, while a shallow network with limited capacity can not deal with such severe degradation. To handle this, we propose a novel Attentive Fractal Network (AFN) to efficiently remove moiré patterns while preserving the texture. With our network  $G(\cdot)$ , the moiré pattern removal task can be formulated as:

$$I_D = G(I_M), \quad (1)$$

where  $I_M$  is the observed image with moiré patterns, and  $I_D$  is our restored clean image.

The proposed AFN adopts a fractal network architecture. With the help of shortcuts and residuals of different recursion levels, the whole network gains the ability to utilize both local and global features for moiré pattern removal. The framework, as shown in Fig. 2 (c), consists of three main parts, which are encoding layers  $G_E$ , an Attentive Fractal Block (AFB)  $G_{\text{AFB}}$ , and decoding layers  $G_D$ .

The encoding layers first transform the input image  $I_M$  into a multi-channel feature  $f_r^0$ , which is the input of the following AFB of recursion level  $r$ .

$$f_r^0 = G_E(I_M). \quad (2)$$

The AFB performs major refinement for the encoded features to obtain  $f_r$ .

$$f_r = G_{\text{AFB}_r}(f_r^0). \quad (3)$$

The feature  $f_r$  is then given to the decoding layers to reconstruct a three-channel clean image  $I_D$ . A global residual connection is added to stabilize the network.

$$I_D = G_D(f_r) + I_M. \quad (4)$$

We also adopt a two-stage augmented refinement strategy to push the ability of AFN further. Specifically, the second stage uses a similar but shallower AFN network with lower recursion levels of AFB to refine the output of the first stage. The details are presented in Sec. 3.3.

### 3.2. Attentive Fractal Block

The proposed AFB is built in a fractal way, and it has a self-similar structure. Each high-level AFB can be constructed with AFBs of lower-level recursively until the level reaches zero. Specifically, as shown in Fig. 2(b), an AFB of recursion level  $s$  consists of  $n_s$  AFBs of recursion level  $s - 1$  as well as a Fusion Unit. In Fig. 2(a), a level 0 AFB is illustrated, and it has a residual shortcut and  $m$  convolution layers followed by LeakyReLU layers. For level  $s$ , the encoded features  $f_s^0$  are fed into the  $n_s$  AFBs of level  $s - 1$  sequentially to obtain  $n_s$  refined features, which are  $f_s^1, \dots, f_s^{n_s}$  respectively:

$$f_s^i = G_{\text{AFB}_{s-1}^{i-1}}(f_s^{i-1}), (1 \leq i \leq n_s). \quad (5)$$

We then adopt a progressive feature fusion strategy, which is to progressively feed-forward the intermediate features of early stages to the end of the current block and fusion

them there. By concatenating the features of the same level, their rich information from different stages help the network learn more thoroughly:

$$f_s^C = [f_s^1, f_s^2, \dots, f_s^{n_s}]. \quad (6)$$

Too many channels might confuse the network with abundant information, so we choose to adopt channel-wise attention with the help of a Squeeze-and-Excitation Layer [9]. By multiplying the assigned learnable weights, the output feature maps are re-weighted explicitly according to their properties.

$$f_s^A = G_{SE_s}(f_s^C) \cdot f_s^C. \quad (7)$$

Finally, we send the features to a Fusion Unit (FU) to narrow down their channels. A local residual connection is also adopted to stabilize the network.

$$f_s = G_{FU_s}(f_s^A) + f_s^0. \quad (8)$$

For level 0 AFB, it sends the input  $f_0^0$  sequentially to  $m$  convolution layers followed by LeakyReLU layers.

$$f_0^i = F_{Leaky}(G_{Conv_i}(f_0^{i-1})), (1 \leq i \leq m). \quad (9)$$

A local shortcut is also adopted to relieve the burden of the network.

$$f_0 = f_0^m + f_0^0. \quad (10)$$

As illustrated above, a high-level AFB is made up of several lower-level AFBs. So actually the input of a level  $s-1$  AFB  $f_{s-1}^0$  is also the input  $f_s^i$  of an intermediate layer in a level  $s$  AFB. And the output of a level  $s-1$  AFB  $f_{s-1}^1$  is also the output  $f_s^{i+1}$  of an intermediate layer in a level  $s$  AFB.

With the help of the fractal structure, the proposed AFB performs refinement operations at different levels. Shortcuts and residual connections effectively solve the vanishing gradients and help the network learn more key features. The progressive fusion of intermediate features boosts the network with abundant information, while the adopted attention mechanism guides the network to focus on essential features. Residual in Residual structures [43] are also implicitly formed, and thus the network is able to extract features of both global structure and high-frequency details for reconstruction. By tweaking the hyper-parameters  $n$ ,  $m$  and  $s$ , the scale and depth of the network can be adjusted neatly for a variety of situations. AFN has provided us with new inspirations towards network architecture design against complex restoration tasks.

### 3.3. Two-Stage Augmented Refinement

Although a single AFN obtains ideal results, we observe that adopting a two-stage strategy further boosts the performance when faced with extremely distorted input images. Specifically, we use a similar but shallower AFN network

as Stage II, and the AFB of Stage II are of lower levels compared to those in Stage I.

We also notice that when flipping the input images, the single AFN can sometimes obtain better results compared to the result when given original input. Intuitively, the result of augmented input alleviates the artifact and provides the second stage with the opportunity to improve the overall restoration quality. Thus, we propose to adopt more transformations for our two-stage restoration strategy. First, we augment the input image using flip, transpose and rotate operations. Then we feed the augmented images into our Stage I AFN network respectively to generate their coarse results. After that, we perform corresponding inverse transforms on the output images to align their directions with the original input. Finally, the coarse results are concatenated together and fed into Stage II AFN for further refinement. Stage II AFN takes the coarse results as input and then learns to selectively fuse them into a clearer image.

The effectiveness of two-stage augmented refinement is illustrated in detail in Sec. 4.6.

## 4. Experiments

### 4.1. Datasets

Previous studies on demoiréing [30, 8] use the DMCNN dataset proposed in [30]. However, its images are captured from screens with random distance and angles, so they require additional alignment via post-processing algorithms. Thus the dataset is limited to moiré patterns of screens only and can not generalize well on other objects with high-frequency textures in everyday life. Additionally, the images in [30] suffer from out-of-focus and pixel density limitation of screens, which lead to color inconsistency and detail loss. And the networks trained on this dataset are required to handle the moiré patterns and the information loss as well to obtain a visually pleasing result. This makes the comparison on this dataset not solely about demoiréing ability. Therefore, we mainly conduct experiments on the CFAMoiré [39] dataset and only provide preliminary results for the DMCNN dataset [30]. In NTIRE20 demoiréing competition, CFAMoiré dataset [39] is proposed to provide a fair comparison of demoiréing ability. It consists of 11,000 image pairs, which are mainly real-world images with high-frequency textured objects. Among them, 500 pairs are left out for test usage, and another 500 pairs are used for the validation phase of the NTIRE20 challenge. We randomly split the released 10,000 pairs into two parts, 9,500 pairs for training and 500 pairs for validation.

In the following sections, we first illustrate the implementation details of our network and introduce the baseline methods. Then we conduct quantitative and qualitative evaluation as well as the ablation study on the CFAMoiré [39] dataset. At last, we provide comparison on the DMCNN dataset [30] and NTIRE20 challenge results.

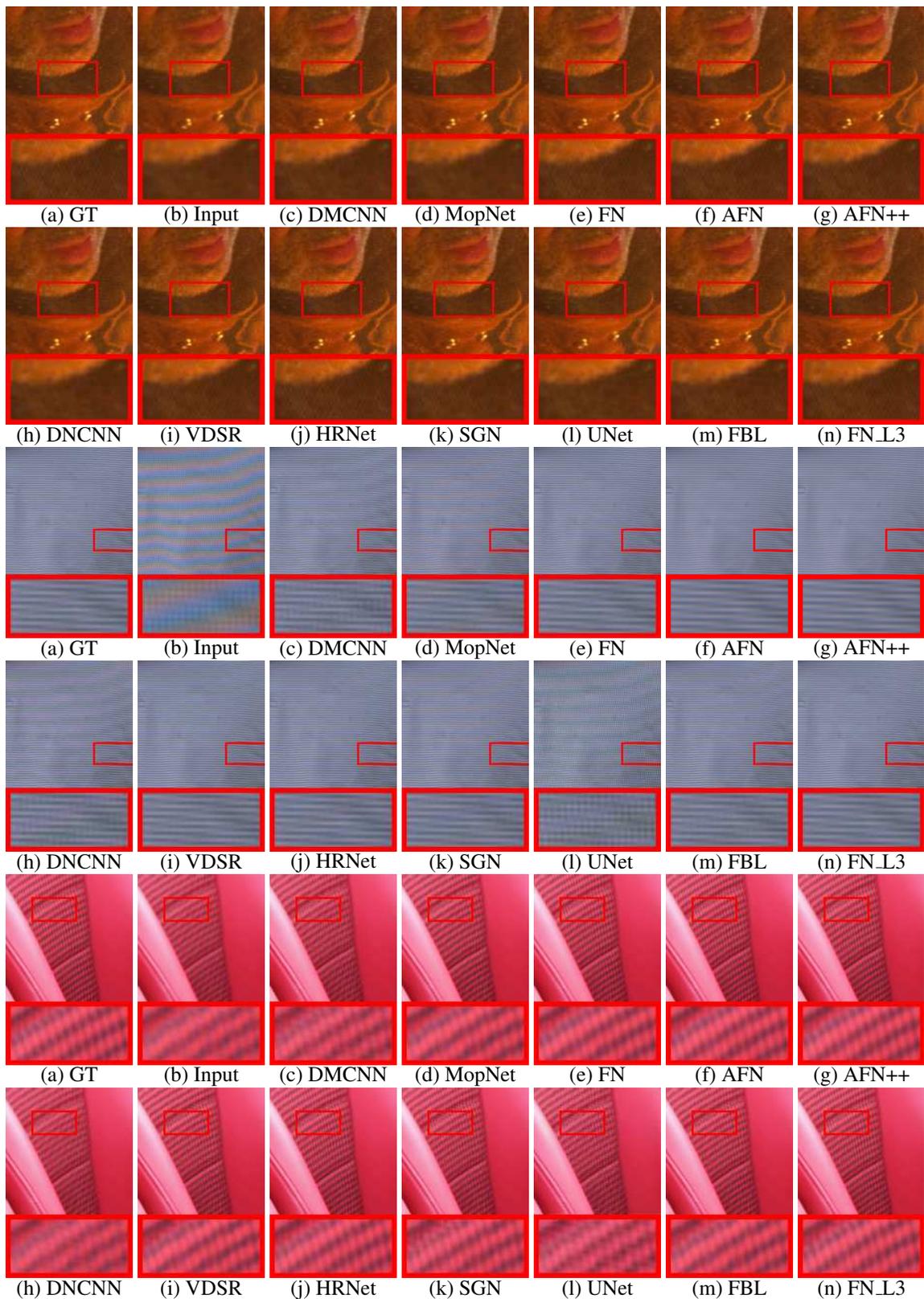


Figure 3: Visual examples of restored results of various methods. Zoom in to view details.

Method	UNet	DNCNN	DMCNN	FBL	MopNet	SGN	VDSR	HRNet	AFN	AFN++
PSNR	30.59	32.91	36.38	37.33	37.45	37.83	37.84	38.50	41.70	42.06
SSIM	0.9410	0.9685	0.9848	0.9873	0.9856	0.9883	0.9883	0.9908	0.9945	0.9945

Table 1: Quantitative comparison on validation set of CFAMoiré [39] dataset. Red and blue indicate the best and the second best performance, respectively. AFN++ denotes the AFN with two-stage augmented refinement. Note that the evaluated method is a 10-layer VDSR instead of the original 20-layer version because the 20-layer version doesn’t converge.

## 4.2. Implementation Details

$n_1, n_2, n_3, n_4, m$  and  $r$  are all set as 4 for Stage I AFN. And  $n_1, n_2, n_3, m$  and  $r$  are all set as 3 for Stage II AFN. The channel size in each  $AFB_0$  is set to 128. We use 2 convolution layers in the encoding and decoding layers. We use 1 convolution layer as the Fusion Unit. The CFAMoiré [39] dataset contains images of size  $128 \times 128$ , and we augment the training data with random flips,  $90^\circ$  rotations and RandomResizedCrop from albumentations [3]. Mini-batch size is set to 16. We use Adam optimizer [13] and set  $\beta_1 = 0.9$  and  $\beta_2 = 0.999$ . The learning rate is initialized as  $1 \times 10^{-4}$  and we adopt a Cosine Annealing strategy [20]. In Stage I, we train the whole network for 200 epochs. Then in Stage II, we fix the Stage I AFN for faster convergence and train Stage II AFN for 50 epochs. For both stages, we train with  $\mathcal{L}_1$  loss only, which minimizes the distance between output and ground truth images. We implement our AFN with PyTorch [26] framework and train them using 4 GeForce RTX 2080 Ti GPUs. Our AFN has 47.7M parameters, and requires 760G MACs for an input image of size  $128 \times 128$ .

## 4.3. Baseline Methods

We compare our method with eight state-of-the-art methods: DMCNN [30], MopNet [8], DNCNN [42], VDSR [12], HRNet [34], SGN [7], UNet [29] and FBL [38]. Among them, DMCNN and MopNet are specially designed methods for the moiré removal task. UNet and HRNet are widely used image restoration architectures. SGN is proposed for the image denoising task. VDSR is originally designed for the single image super resolution task. FBL is a network proposed for rain streaks removal. During the evaluation, we train the compared models from scratch except for MopNet. Specifically for MopNet, we load the pre-trained checkpoint released by the author and remove its classification branch, because that branch requires additional labels other than the target image. After that, we fine-tune the whole network until convergence. All methods except DMCNN and VDSR are directly evaluated with online available codes, while DMCNN is implemented by ourselves and the architecture of VDSR is modified from 20 layers to 10 layers because the original 20-layer version doesn’t converge. Two metrics, Peak Signal-to-Noise Ratio (PSNR) [10] and Structure Similarity Index (SSIM) [35], are adopted, and we evaluate the results in all RGB channels.

Methods	$r$	SE	Stage II	PSNR	SSIM
AFN++	4	✓	✓	42.06	0.9945
AFN	4	✓		41.70	0.9945
FN	4			40.73	0.9932
FN L3	3			40.01	0.9922

Table 2: Ablation of our AFN.

## 4.4. Quantitative evaluation

In this section, we provide quantitative evaluation of our proposed AFN and state-of-the-art methods on the validation set of CFAMoiré [39] dataset. As shown in Tab. 1, we can observe that some widely used image restoration backbones, namely SGN and HRNet, obtain pleasing results compared to the previous moiré pattern removal networks. This is mainly because MopNet and DMCNN are specially designed for the complex degradations in the DMCNN dataset [30], and thus generate blurry results overall instead of realistic textures. In comparison, as image restoration frameworks for super resolution and noise removal, SGN, VDSR and HRNet adopt fine designs for detail preserving reconstruction. It can also be seen that our method outperforms other state-of-the-art methods by 3.2dB. With the proposed two-stage augmented refinement strategy, noted as AFN++, we further obtain a large gain of 0.36dB.

## 4.5. Qualitative evaluation

The visual examples of qualitative evaluation are provided in Fig. 3. As emphasized in the red boxes, previous methods show limited effectiveness, and our proposed AFN obtains cleaner results with less artifact. Specifically, the first input image contains stripe shaped artifacts, and it can be seen that similar patterns still exist on the outputs of DNCNN, DMCNN and MopNet. The second input image is severely contaminated, and the results of DNCNN, SGN and UNet are over-smoothed with unpleasant artifacts. The third group contains an image with high-frequency stripe-shaped textures. From the last two rows, it can be observed that our AFN restores its details while other methods, such as FBL, VDSR and HRNet, blur the intrinsic patterns.

## 4.6. Ablation Studies

In this section, we conduct ablation experiments on parts of our model for comparison. The visual results are shown

in Fig. 3 and the quantitative results are provided in Tab. 2. AFN++ denotes the proposed AFN with Stage II augmented refinement. AFN means that only Stage I is adopted. FN means we remove the attention mechanism inside each AFB. FN L3 has a similar structure with FN but it contains top-level AFB of level 3. As listed in the table and provided in the figure, the proposed AFN++ generates the best restored results. The comparison of the variants proves that the proposed designs all contribute to the performance gain of the restored images. The comparison between AFN++ and AFN shows the effectiveness of Stage II augmented refinement. The comparison between AFN and FN demonstrates the contribution of channel-wise attention mechanism. And the comparisons between FN and FN L3 prove the importance of the fractal architecture.

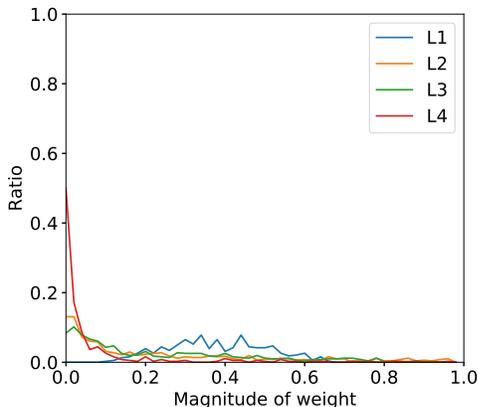


Figure 4: Visualization of the channel-wise weight generated by the attention mechanism. The L1 to L4 in the legend indicate the corresponding levels of the blocks. For simplicity, we provide the results of the last SE layer of each level.

We also investigate the effectiveness of the Squeeze-and-Excitation attention mechanism. Specifically, we obtain the channel-wise weight generated by AFN when given Fig. 1 (a) as the input image. As shown in Fig. 4, the weights vary among the channels. It can be inferred that the attention mechanism contributes to better restoration result by guiding the network to focus on more important feature maps.

#### 4.7. Evaluation in NTIRE20 Challenge

In this section, we provide the results on the unreleased test set of NTIRE20 Challenge. Specifically, we submitted our models on 3 different tracks, including single image demoiréing, burst demoiréing and single image deblurring. Quantitative results are presented in Tab. 3, 4, and 5. Our AFN wins the burst demoiréing track and achieves second place in single image demoiréing and single image deblurring tracks. Note that the ability of deblurring further proves that our proposed architecture is especially good at preserving details and reconstructing textures.

Methods	PSNR	SSIM
AFN (Ours)	<b>41.95</b>	<b>0.99</b>
2nd method	41.88	0.99
3rd method	40.64	0.99
4th method	40.33	0.99
5th method	39.05	0.99

Table 3: Quantitative results on NTIRE20 Challenge Burst Demoiréing track [39]. Red and blue indicate the best and the second best performance, respectively. Due to the time limit of the challenge, we only used one of the frames for restoration. We believe that our performance will be further boosted with the help of the additional six frames.

Methods	PSNR	SSIM
1st method	<b>42.14</b>	<b>0.99</b>
AFN (Ours)	41.95	0.99
3rd method	41.84	0.99
4th method	41.11	0.99
5th method	41.04	0.99

Table 4: Quantitative results on NTIRE20 Challenge Single Image Demoiréing track [39]. Red and blue indicate the best and the second best performance, respectively.

Methods	PSNR	SSIM
1st method	<b>34.44</b>	<b>0.9412</b>
AFN (Ours)	34.20	0.9392
3rd method	33.35	0.9283
4th method	33.07	0.9242
5th method	32.61	0.9198

Table 5: Quantitative results on NTIRE20 Challenge Deblur track [25]. Red and blue indicate the best and the second best performance, respectively.

Methods	PSNR	SSIM
MopNet-Full	<b>27.75</b>	<b>0.89</b>
MopNet-w/o Class Label	27.44	0.89
MopNet-w/o Edge	27.07	0.88
AFN L3 (Ours)	27.00	0.83
MopNet-w/o Edge & Class Label	26.62	0.88
DMCNN	26.77	0.871
UNet	26.49	0.864
VDSR	24.68	0.837
DNCNN	24.54	0.834

Table 6: Quantitative results on DMCNN Dataset. Red and blue indicate the best and the second best performance, respectively.

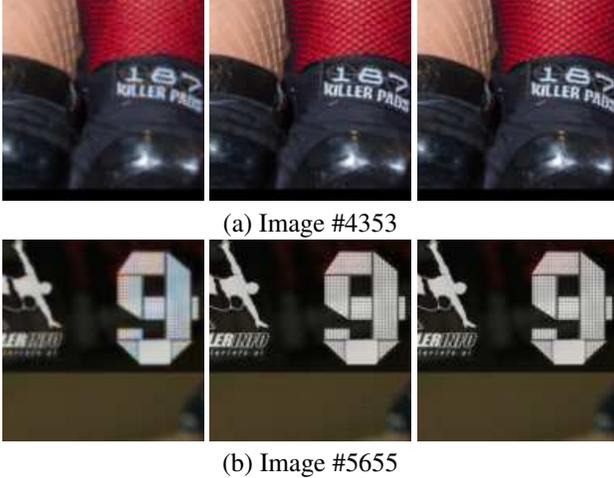


Figure 5: Example images containing moiré pattern contaminated texts. Images in each row are input, result from AFN++, and ground truth, respectively.

#### 4.8. Evaluation in the DMCNN Dataset

In this section, we provide our preliminary results in the DMCNN Dataset [30]. As illustrated above, the images in this dataset suffer from detail loss caused by out-of-focus and pixel density limitation of screens. We believe it’s not an ideal option for comparison of demoiré ability. However, it can be considered as a metric of general restoration of complex degradations. As shown in Tab. 6, our AFN of level 3 achieves comparable performance as the version of MopNet without the guidance of edge and class label.

### 5. Discussions

#### 5.1. Application

It has attracted increasing attention [27, 11, 19] to adopt enhancement methods as pre-processing for downstream high-level computer vision tasks. In order to investigate the impact of moiré patterns on OCR performance, we employ Google Cloud Vision API [1] to test whether the OCR results are improved on several hard test cases. The results are shown in Tab. 7 and Fig. 5. As observed, removing moiré patterns with our method as a pre-processing step brings significant improvement to the OCR accuracy. As suggested in [16], the high-level task performance serves as a fixed semantic-aware metric, and can be seen as an indirect approval of semantic information preservation of our method.

#### 5.2. Limitation and Future Work

We observe dataset bias between DMCNN dataset and CFAMoiré [39] dataset. Specifically, a model trained on DMCNN dataset cannot produce ideal results on CFAMoiré [39] dataset, and a model trained on CFAMoiré [39] dataset cannot generate pleasing images on

Image ID	#4353	#5655
Input	KRLER PA	-
AFN++	KILLER PAS	TERINFO Teriarel
GT	KILLER PA	TERINFO Teriatel

Table 7: OCR results yielded by Google Cloud Vision API. Numbers after sharp symbols are the indices in CFAMoiré [39] training set. Dash symbol indicates that Vision API detects no text in the given image.

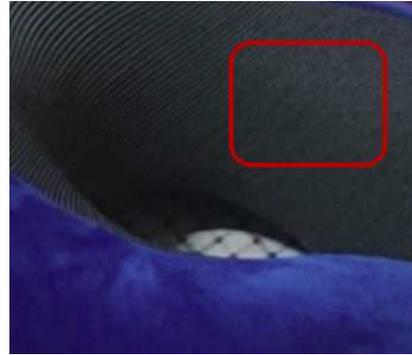


Figure 6: An example of hard case when the moiré patterns (in the red box) appear similar to the original textures.

DMCNN dataset. We believe that the generalizability of moiré removal methods is worth further investigating, considering that the moiré patterns in the wild cover a large variety of shapes and colors.

Current datasets mainly contain cases when a whole image is contaminated, however, in real life photography, it is common case that only parts of an image are affected by the moiré artifacts. The moiré removal networks sometimes cannot accurately identify the moiré patterns and fail to estimate appealing results. An example is provided in Fig. 6, the network mistakenly take the moiré patterns as the textures and chooses to leave the artifacts untouched. We will try to improve such cases in our future work.

### 6. Conclusion

In this paper, we introduce our solution in NTIRE20 image restoration and enhancement challenges. We propose Attentive Fractal Network, a fractally stacked network to effectively restore moiré contaminated images. First, we build a fractal network with progressive feature fusion and channel-wise attention guidance. Then we adopt a two-stage augmented refinement strategy to further boost the performance. Extensive experiments demonstrate that our method outperforms existing state-of-the-art methods. Our AFN also wins the burst demoiré track and achieves second place in single image demoiré and single image deblurring tracks in NTIRE20 Challenges.

## References

- [1] Google Cloud Vision API. <https://cloud.google.com/vision/>. 8
- [2] Sattva Descreen. <http://www.descreen.net/eng/soft/descreen/descreen.htm>. 1, 2
- [3] Buslaev A, Parinov A, Khvedchenya E, Iglovikov V.I, and A. A. Kalinin. Albumentations: fast and flexible image augmentations. *ArXiv e-prints*, 2018. 6
- [4] Isaac Amidror. Sub-nyquist artefacts and sampling moiré effects. *Royal Society open science*, 2(3):140550, 2015. 1
- [5] Bryce E Bayer. Color imaging array, July 20 1976. US Patent 3,971,065. 1, 2
- [6] Lars A Bergkvist and Ivan Forsen. Leading mark indicator, December 16 1986. US Patent 4,629,325. 1
- [7] Shuhang Gu, Yawei Li, Luc Van Gool, and Radu Timofte. Self-guided network for fast image denoising. In *Proc. International Conference on Computer Vision*, 2019. 2, 6
- [8] Bin He, Ce Wang, Boxin Shi, and Ling-Yu Duan. Mop moire patterns using mopnet. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2424–2432, 2019. 1, 2, 4, 6
- [9] Jie Hu, Li Shen, and Gang Sun. Squeeze-and-excitation networks. In *Proc. International Conference on Computer Vision and Pattern Recognition*, 2018. 4
- [10] Quan Huynh-Thu and Mohammed Ghanbari. Scope of validity of psnr in image/video quality assessment. *Electronics letters*, 44(13):800–801, 2008. 6
- [11] Yifan Jiang, Xinyu Gong, Ding Liu, Yu Cheng, Chen Fang, Xiaohui Shen, Jianchao Yang, Pan Zhou, and Zhangyang Wang. Enlightengan: Deep light enhancement without paired supervision. *arXiv preprint arXiv:1906.06972*, 2019. 8
- [12] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee. Accurate image super-resolution using very deep convolutional networks. In *Proc. International Conference on Computer Vision and Pattern Recognition*, 2016. 6
- [13] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 6
- [14] Junhyung Kwak and Donghee Son. Fractal residual network and solutions for real super-resolution. In *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2019. 2
- [15] Gustav Larsson, Michael Maire, and Gregory Shakhnarovich. Fractalnet: Ultra-deep neural networks without residuals. *arXiv preprint arXiv:1605.07648*, 2016. 2
- [16] Boyi Li, Wenqi Ren, Dengpan Fu, Dacheng Tao, Dan Feng, Wenjun Zeng, and Zhangyang Wang. Benchmarking single-image dehazing and beyond. *IEEE Transactions on Image Processing*, 28(1):492–505, 2018. 8
- [17] Ruoteng Li, Loong-Fah Cheong, and Robby T Tan. Heavy rain image restoration: Integrating physics model and conditional adversarial learning. In *Proc. International Conference on Computer Vision and Pattern Recognition*, 2019. 2
- [18] Fanglei Liu, Jingyu Yang, and Huanjing Yue. Moiré pattern removal from texture images via low-rank and sparse matrix decomposition. In *2015 Visual Communications and Image Processing (VCIP)*, pages 1–4. IEEE, 2015. 1, 2
- [19] Yu Liu, Guanlong Zhao, Boyuan Gong, Yang Li, Ritu Raj, Niraj Goel, Satya Kesav, Sandeep Gottimukkala, Zhangyang Wang, Wenqi Ren, et al. Improved techniques for learning to dehaze and beyond: A collective study. *arXiv preprint arXiv:1807.00202*, 2018. 8
- [20] Ilya Loshchilov and Frank Hutter. Sgdr: Stochastic gradient descent with warm restarts. *arXiv preprint arXiv:1608.03983*, 2016. 6
- [21] Richard F Lyon and Paul M Hubel. Eyeing the camera: Into the next century. In *Color and Imaging Conference*, volume 2002, pages 349–355. Society for Imaging Science and Technology, 2002. 1, 2
- [22] Daiqian Ma, Renjie Wan, Boxin Shi, Alex C Kot, and Ling-Yu Duan. Learning to jointly generate and separate reflections. In *Proc. International Conference on Computer Vision and Pattern Recognition*, 2019. 2
- [23] Richard Billings Merrill. Color separation in an active pixel cell imaging array using a triple-well structure, October 12 1999. US Patent 5,965,875. 1, 2
- [24] Shinichi Mihara, Hirokazu Konishi, Toyoharu Hanzawa, Masahito Watanabe, Atsujiro Ishii, Tetsuhide Takeyama, and Ayami Imamura. Electronic image pickup system, October 14 2008. US Patent 7,436,599. 1, 2
- [25] Seungjun Nah, Sanghyun Son, Radu Timofte, and Kyoung Mu Lee. Ntire 2020 challenge on image and video deblurring. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2020. 7
- [26] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. Pytorch: An imperative style, high-performance deep learning library. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d Alche-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems 32*, pages 8024–8035. Curran Associates, Inc., 2019. 6
- [27] Rui Qian, Robby T Tan, Wenhan Yang, Jiajun Su, and Jiaying Liu. Attentive generative adversarial network for rain-drop removal from a single image. In *Proc. International Conference on Computer Vision and Pattern Recognition*, pages 2482–2491, 2018. 8
- [28] Yanyun Qu, Yizi Chen, Jingying Huang, and Yuan Xie. Enhanced pix2pix dehazing network. In *Proc. International Conference on Computer Vision and Pattern Recognition*, 2019. 2
- [29] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015. 6
- [30] Yujing Sun, Yizhou Yu, and Wenping Wang. Moiré photo restoration using multiresolution convolutional neu-

- ral networks. *IEEE Transactions on Image Processing*, 27(8):4160–4172, 2018. 1, 2, 4, 6, 8
- [31] Seiji Tanaka. Color imaging apparatus, September 10 2013. US Patent 8,531,563. 1, 2
- [32] Pericles S Theocaris. *Moiré fringes in strain analysis*. Elsevier, 1969. 1
- [33] Richard M Turner and Rudolph J Guttosch. Development challenges of a new image capture technology: foveon x3 image sensors. In *International congress of imaging science*, pages 175–181, 2006. 1, 2
- [34] Jingdong Wang, Ke Sun, Tianheng Cheng, Borui Jiang, Chaorui Deng, Yang Zhao, Dong Liu, Yadong Mu, Mingkui Tan, Xinggang Wang, et al. Deep high-resolution representation learning for visual recognition. *arXiv preprint arXiv:1908.07919*, 2019. 6
- [35] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004. 6
- [36] Chen Wei\*, Wenjing Wang\*, Wenhan Yang, and Jiaying Liu. Deep retinex decomposition for low-light enhancement. In *Proc. British Machine Vision Conference*, 2018. 2
- [37] Zhouping Wei, Jian Wang, Helen Nichol, Sheldon Wiebe, and Dean Chapman. A median-gaussian filtering framework for moiré pattern noise removal from x-ray microscopy image. *Micron*, 43(2-3):170–176, 2012. 1, 2
- [38] Wenhan Yang, Shiqi Wang, Dejia Xu, Xiaodong Wang, and Jiaying Liu. Towards scale-free rain streak removal via self-supervised fractal band learning. In *Proc. AAAI Conference on Artificial Intelligence*, 2020. 2, 6
- [39] Shanxin Yuan, Radu Timofte, Ales Leonardis, Gregory Slabaugh, et al. Ntire 2020 challenge on image demoiréing: Methods and results. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2020. 4, 6, 7, 8
- [40] Shanxin Yuan, Radu Timofte, Gregory Slabaugh, Ales Leonardis, Bolun Zheng, Xin Ye, Xiang Tian, Yaowu Chen, Xi Cheng, Zhenyong Fu, Jian Yang, Ming Hong, Wenying Lin, Wenjin Yang, Yanyun Qu, Hong-Kyu Shin, Joon-Yeon Kim, Sung-Jea Ko, Hang Dong, Yu Guo, Jie Wang, Xuan Ding, Zongyan Han, Sourya Dipta Das, Kuldeep Purohit, Praveen Kandula, Maitreya Suin, and A. N. Rajagopalan. Aim 2019 challenge on image demoiréing: Methods and results. *arXiv preprint arXiv:1911.03461*, 2019. 2
- [41] Hongguang Zhang, Yuchao Dai, Hongdong Li, and Piotr Koniusz. Deep stacked hierarchical multi-patch network for image deblurring. In *Proc. International Conference on Computer Vision and Pattern Recognition*, pages 5978–5986, 2019. 2
- [42] Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, and Lei Zhang. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE Transactions on Image Processing*, 26(7):3142–3155, 2017. 6
- [43] Yulun Zhang, Kunpeng Li, Kai Li, Lichen Wang, Bineng Zhong, and Yun Fu. Image super-resolution using very deep residual channel attention networks. In *Proc. European Conference on Computer Vision*, 2018. 2, 4