NTIRE 2020 Challenge on Image Demoireing: Methods and Results

Shanxin Yuan   Radu Timofte   Aleš Leonardis   Gregory Slabaugh

Xiaotong Luo   Jiangtao Zhang   Yanyun Qu   Ming Hong   Yuan Xie   Cuihua Li   Dejia Xu   Yihao Chu

Qingyan Sun   Shuai Liu   Ziyao Zong   Nan Nan   Chenghua Li   Sangmin Kim   Hyungjoon Nam

Jisu Kim   Jechang Jeong   Manri Cheon   Sung-Jun Yoon   Byungyeon Kang   Junwoo Lee   Bolun Zheng

Xiaohong Liu   Linhui Dai   Jun Chen   Xi Cheng   Zhenyong Fu   Jian Yang   Chul Lee   An Gia Vien

Hyunkook Park   Sabari Nathan   M.Parisa Beham   S Mohamed Mansoor Roomi   Florian Lemarchand

Maxime Pelcat   Erwan Nogues   Densen Puthussery   Hrishikesh P S   Jiji C V   Ashish Sinha   Xuan Zhao

Abstract

This paper reviews the Challenge on Image Demoireing that was part of the New Trends in Image Restoration and Enhancement (NTIRE) workshop, held in conjunction with CVPR 2020. Demoireing is a difficult task of removing moire patterns from an image to reveal an underlying clean image. The challenge was divided into two tracks. Track 1 targeted the single image *demoireing problem, which seeks to remove moire patterns from a single image. Track 2 focused on the* burst *demoireing problem, where a set of degraded moire images of the same scene were provided as input, with the goal of producing a single demoired image as output. The methods were ranked in terms of their fidelity, measured using the peak signal-to-noise ratio (PSNR) between the ground truth clean images and the restored images produced by the participants' methods. The tracks had 142 and 99 registered participants, respectively, with a total of 14 and 6 submissions in the final testing stage. The entries span the current state-of-the-art in* image and burst image demoireing *problems.*

1. Introduction

Digital photography has matured in the past years with the new advancements in relevant research areas, including image denoising [61, 19, 2], image demosaicing [16], super-resolution [37, 46, 47, 8, 9, 31, 23], deblurring [33], dehazing [4], quality mapping [22], automatic white balance [7], and high dynamic range compression [17]. Moire

aliasing is a less addressed and fundamental problem. In the case of digital photography, moire aliasing occurs when the camera's color filter array (CFA) interferes with high frequency scene content close to the resolution of CFA grid. The high frequency regions result in undersampling on the sensor color filter array (CFA) and when demosaiced, can create disruptive colorful patterns that degrade the image. For example, moire aliasing is likely to happen when taking pictures of clothing or long-distance building's tiles.

The AIM19 demoireing challenge [59] addressed a more specific scenario of photography of digital screens. In this scenario, moire patterns appear when the CFA interferes with the LCD screen's subpixel layout. However, in the photography natural scenes, moire aliasing is also a less addressed problem, where the CFA interferes with high frequency scene content. Our NTIRE2020 demoireing challenge addresses this more general case.

By engaging the academic community with this challenging image enhancement problem, a variety of methods have been proposed, evaluated, and compared using a common dataset. In this paper, we also propose a new dataset called *CFAMoire* consisting of 11,000 image pairs (image with moire patterns in the high frequency areas, and clean ground truth).

This challenge is one of the NTIRE 2020 associated challenges on: deblurring [34], nonhomogeneous dehazing [5], perceptual extreme super-resolution [60], video quality mapping [15], real image denoising [1], real-world super-resolution [32], spectral reconstruction from RGB image [6] and demoireing.

2. The Challenge

The demoireing challenge was hosted jointly with the *New Trends in Image Restoration and Enhancement (NTIRE)* workshop held in conjunction with the IEEE Computer Society Conference on Computer Vision and Pattern

Recognition (CVPR) 2020.

The task of image demoireing is to design an algorithm to remove moire patterns from an input image. To help build algorithms, particularly those based on machine learning, a novel dataset was produced for the challenge.

2.1. CFAMoire Dataset

As an essential step towards reducing moire effects, we proposed a novel dataset, called *CFAMoire*. It consists of 10,000 training, 500 validation, and 500 test images with 128×128 resolution. The images are high quality in terms of the reference frames and different moire patterns. It also covers balanced content, including clothing and buildings, where the high frequency repetitive patterns interfere with the camera's color filter array.

The clean images are sampled or cropped from existing datasets, including DeepJoint [16], HDR+ [20], and DIV8K [18]. The corresponding moire images are generated through a remosaicing and demosaicing process, where moire artifacts will appear in the high frequency area. To collect the clean images, we followed a two-step strategy through first automatically selecting images by quantification of moire artifacts in the Fourier domain, followed by manual selection. The automatic selection step is conducted by measuring the frequency change between the clean image and the demosaiced image [16].

Assuming the clean image and the demosaiced image to be I_c and I_m. Both images are first converted to the Lab space and then a 2D Fourier transform is applied to each channel to get $F(I_c)$ and $F(I_m)$. The frequency ratio is calculated by

$$\rho(w) = \begin{cases} log(\frac{|F(I_m)|^2+\eta}{|F(I_c)|^2+\eta}) & if \quad |w| < \gamma \\ 1 & otherwise \end{cases} \quad (1)$$

where w is the spatial frequency and γ is a threshold. Following the setting of [16], we only compare the ratio in frequencies lower than γ, which is set to 0.95π, to mitigate high-frequency noise. The ratio map is then smoothed with Gaussian blur. If the maximum ratio value across all channels and frequencies exceeds a threshold of 2, then the image is selected.

This technique can produce a large number of candidate image pairs for the dataset. However, it cannot detect clean images that are already corrupted with moire artifacts. Moreover, since it works in the Fourier domain and requires some manually set parameters, it may select suboptimal image pairs. Therefore, the organizers spent one week on manual selection. As a result, only 10% of the pairs from the automatic step are selected. Figure 1 shows some examples.

2.2. Tracks and Evaluation

The challenge had two tracks: **Single image** track and **Burst** track.

Single image: In this track, participants developed demoire methods to achieve a moire-free image with the best fidelity compared to the ground truth. Figure 1 shows some image pairs.

Burst: Similar to the Single image track, this track works on a burst of images. To generate the input data, apart from the reference moire image, 6 supporting images are generated by first applying random homographies to the clean image and then by going through the mosaicing and demosaicing step, see Figure 2 for examples.

For both tracks, we used the standard Peak Signal To Noise Ratio (PSNR) measure to rank methods. Additionally, the Structural Similarity (SSIM) index was computed. Implementations of PSNR and SSIM are found in most image processing toolboxes. For each method we report the average results over all the processed images.

2.3. Competition

Platform: The CodaLab platform was used for this competition. To access the data and submit their demoired image results to the CodaLab evaluation server each participant had to register.

Challenge phases: (1) Development (training) phase: the participants received both moire and moire-free training images of the dataset. (2) Validation phase: the participants had the opportunity to test their solutions on the moire validation images and to receive immediate feedback by uploading their results to the server. A validation leaderboard was available. (3) Final evaluation phase: after the participants received the moire test images and clean validation images, they submitted their source code / executable, demoired images, and a factsheet describing their method before the challenge deadline. One week later, the final results were made available to the participants.

3. Results

For the single image track, there were 142 registered teams, and 14 teams entered the the final phase by submitting results. Table 1 reports the final test results, rankings of the challenge, self-reported runtimes and major details from the factsheets. The leading entry was from the HaiYun team, scoring a PSNR of 42.14 dB. Second and third place entries were by teams OIerM and Alpha respectively.

For the burst track, there were 99 registered teams, and 6 teams entered the final phase by submitting results. Table 2 reports the results and details from the factsheets. The leading entry was from the OIerM team, scoring a PSNR of 41.95. Second and third place entries were by teams Alpha and Mac AI respectively. Please note that the best PSNR

Figure 1: Examples from the *CFAMoire* dataset. Top row are moire images, bottom row are ground truth.
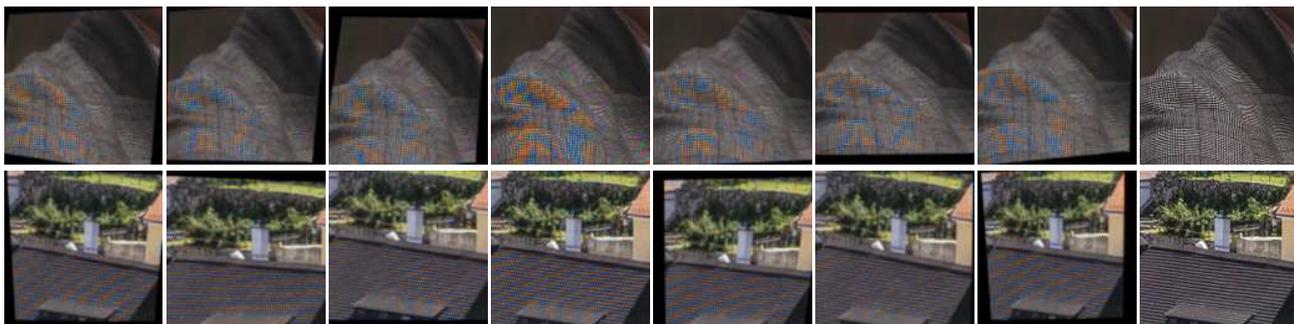
Figure 2: Two examples for the burst image track. Random homographies are applied to the clean image before the remosaicing and demosaicing step. The last column is the ground truth. With slight viewpoint changes, the moire patterns change dramatically. A burst of images can provide extra information to the reference image.

Team	PSNR	SSIM	Time (train, test)	GPU	Platform	Loss	Ensemble	GFlops	#Para
HaiYun	42.14	0.99	3 days, 0.45	GTX Titan XP	Pytorch 1.0.0	L1 (RGB) + L1 (DCT)	Self-ensemble	45.17	144.33M
OIerM	41.95	0.99	67 hours, 0.1	RTX 2080Ti	PyTorch	L1	Refinement x16	-	56.2M
Alpha	41.84	0.99	5 days, 0.34	RTX 2080Ti	PyTorch	L1 charbonnier	Self-ensemble x4	267	23M
Reboot	41.11	0.99	10 days, 0.87	RTX 2080Ti	PyTorch 1.4	L1 (RGB) + L1 (UV)	Self-ensemble x4	-	164M
LIPT	41.04	0.99	7 days, 1.5	Tesla V100	TF 2.0	L1 (RGB) + L1 (∇RGB) + L1 (DCT)	Self-ensemble x8	-	5.1M
iipl-hdu	40.68	0.99	3 days, 0.1	2080Ti	Tensorflow 1.14	L1 + ASL	-	-	24M
Mac AI	40.60	0.99	-, 0.1635	GTX 1080Ti	PyTorch 1.1	-	Self-ensemble x8	-	13.6M
MoePhoto	39.05	0.99	24 hours, 0.14	RTX2080Ti (11GB) x4	PyTorch 1.4	L1 charbonnier + L1 Wavelet	None	46.38	6.48M
DGU-CILab	38.28	0.98	3 days, 0.187	RTX 2080Ti	Pytorch 1.3	L2	Self-ensemble x8	-	14M
Image Lab	36.73	0.98	-,0.01944	GPU with 8GB memory	TF 1.10	L2 + SSIM + Sobel	-	-	1.6M
VAADER	36.63	0.98	2days, 0.035	2080Ti train, GTX 1080Ti (test)	PyTorch 1.4	L1	Self-ensemble x8	-	13M
CET_CVLab	32.07	0.95	5 days,0.192	Nvidia Quadro K6000	TF 2.1	L1 + color loss	No	-	-
sinashish	30.69	0.92	6 hours, 0.045	Tesla K80	PyTorch 1.14	L2	No	-	-
NTIREXZ	29.52	0.90	18 hours, 0.115	Nvidia 1060	-	L1		-	-
no processing	25.45	0.77	-	-	-	-		-	-

Table 1: Final results for the **Single image** demoire challenge track (Track 1).

in the burst track is *lower* than that of single frame track, which is counter-intuitive. One would expect the additional frames to provide more data useful to produce a better quality result. However, after the challenge's deadline, Alpha team submitted new results that significantly increased the PSNR to 45.32 (larger than the best result on single frame track) by making use of alignment across the burst.

Looking across entries, some interesting trends were noted. In particular,

- **Ensembles:** Most solutions used self-ensemble [48] x8. Small improvements are reported by the solu-

tions. Besides self-ensemble, a new fusion method is also proposed by OIerM team. OIerM team performed 16 invertible transforms including rotating, flipping, transposing and their combinations on a single image. Then they perform corresponding inverse transforms on the outputs and concatenated them together to get augmented results. Then they train a smaller network to fuse the augmented results. This strategy brings +0.35dB PNSR over the single method on validation set.

- **Multi-scale strategy:** Most solutions adopted a multi-

Team	PSNR	SSIM	Time (train, test)	GPU	Platform	Loss	Ensemble	GFlops	#Para
OIerM	41.95	0.99	67 hours, 0.1	RTX 2080Ti	PyTorch	**Loss**	Refinement x16		56.2M
Alpha	41.88	0.99	3 days, 0.34	RTX 2080Ti	Pytorch	L1 charbonnier	Self-ensemble x8		
Alpha*	45.32	1.00	5 days, 0.34	RTX 2080Ti	PyTorch	L1 charbonnier	Self-ensemble x8		
Mac AI	40.64	0.99	-,0.2753s	GTX 1080Ti	PyTorch 1.1	-	Self-ensemble x8	-	13.8M
Reboot	40.33	0.99	10 days, 0.22	RTX 2080Ti	PyTorch 1.4	L1 (RGB) + L1 (UV)	None	-	21M
MoePhoto	39.05	0.99	24 hours, 0.14	RTX 2080T x4	PyTorch 1.4	L1 charbonnier + L1 Wavelet	None	46.38	6.48M
DGU-CILab	38.50	0.99	4 days, 0.1636	RTX 2080Ti	Pytorch 1.3	-	Self-ensemble x8	-	14M
no processing	25.45	0.77	-	-	-	-	-	-	-

Table 2: Final results for the **Burst** demoire challenge track (Track 2). * submission after the deadline.

scale strategy as a mechanism to handle moire patterns of different frequencies.

- **New loss functions**: Although many solutions choose the traditional L1 loss function. A few new loss functions are used, including coral loss by HaiYun team, L1 Wavelet loss by MoePhoto team, color loss by CET_CVLab.

The next section describes briefly the method of each team, while in the Appendix A the team members and their affiliations are provided.

4. Challenge Methods

4.1. HaiYun team
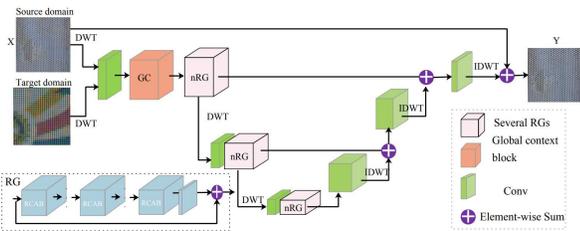
Figure 3: The overall network architecture of the proposed AWUDN.

HaiYun team propose a deep wavelet network with domain adaptation for single image demoireing, dubbed AWUDN [53], as shown in Figure 3. The whole network is an U-Net structure, where the downsampling and upsampling of feature maps are replaced with discrete wavelet transform (DWT) and inverse discrete wavelet transform (IDWT) for reducing computation complexity and information loss. Therefore, the feature mapping is performed in wavelet domain, where the basic block adopts the residual-in-residual structure [62] for extracting more residual information effectively. Considering that the dataset provided by the competition has strong self-similarity, i.e., similar texture structure, the global context block [10] as shown in Figure 4 is introduced in the front of network structure. It can help establish the relationship between two distant pixels to better use the internal information of the image for restoring texture details. Moreover, there may exist slight domain difference between the source domain training data and the
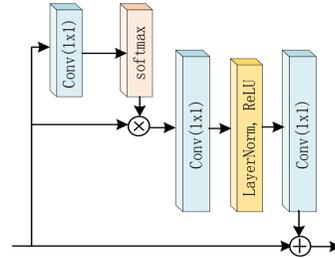
Figure 4: Global context block

target domain testing data. It means that the distribution of moire images in the training set and the moire images in the test set is inconsistent, which can constrain the performance improvement of the model pretrained on the training dataset. CORAL loss [43] gives us some inspiration, which defines the measurement difference of the second-order statistics between features in the source domain and the target domain. Therefore, the pretrained model WUDN is fine-tuned using coral loss for reducing the domain shift of training moire dataset and testing moire dataset in the testing phase. Self-ensemble strategy is adopted to further improve performance. The proposed solution can obtain significant quantitative and qualitative results.

4.2. OIerM team

OIerM team propose an Attentive Fractal Network (AFN) [54] to effectively solve the demoiré problem. First, they construct an Attentive Fractal Block via progressive feature fusion and channel-wise attention guidance, then they stack our AFB in a fractal way inspired by FBL-SS [55]. Finally, to further boost the performance, they adopt a two-stage augmented refinement strategy.

The proposed AFN adopts a fractal network architecture. With the help of shortcuts and residuals of different levels, the whole network gains the ability to utilize both local and global features for moiré pattern removal. The framework, as shown in Fig. 5 (c), consists of three main parts, which are encoding layers G_E, an Attentive Fractal Block (AFB) G_{AFB}, and decoding layers G_D.

The encoding layers first transform the input image I_M into a multi-channel feature f_s^0, which is the input of the
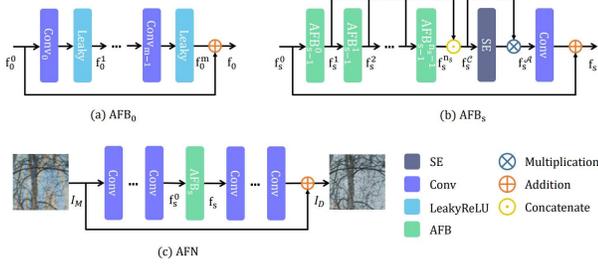
Figure 5: Architecture of AFN. (a) shows the structure of the basic block AFB_0. (b) shows the hierarchical architecture of AFB_s, where s is the recursion level of the block. (c) illustrates the structure of AFN, which consists of an encoding block, an AFB_s, and an decoding block.

following AFB of level s.

$$f_s^0 = G_E(I_M). \tag{2}$$

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

$$f_s = G_{AFB_s}(f_s^0). \tag{3}$$

The feature f_s 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.

$$I_D = G_D(f_s) + I_M. \tag{4}$$

They 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 levels of AFB to refine the output of the first stage.

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. 5(b), an AFB of level s consists of n_s AFBs of level $s-1$ as well as a Fusion Unit. In Fig. 5(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 each of the n_s AFBs of level $s-1$ sequentially to obtain n_s refined features, which are $f_s^1, \cdots, f_s^{n_s}$ respectively:

$$f_s^i = G_{AFB_{s-1}^{i-1}}(f_s^{i-1}), (1 \le i \le n_s). \tag{5}$$

They 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:

$$f_s^C = [f_s^1, f_s^2, \cdots, f_s^{n_s}]. \tag{6}$$

Too many channels might confuse the network with abundant information, so they choose to adopt channel-wise attention with the help of a Squeeze-and-Excitation Layer [21]. By multiplying the assigned learnable weights, the output feature maps are re-weighted explicitly according to their properties.

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

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

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

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

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

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

$$f_0 = f_0^m + f_0^0. \tag{10}$$

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} is also the output f_s^{i+1} of an intermediate layer in a level s AFB.

4.3. Alpha team

Alpha team propose the *MMDM: Multi-frame and Multi-scale for Image Demoireing* [29], see Figure 6. They propose a feature extraction and reconstruction module (FERM) based on RCAN [62] by removing all channel attention modules and upsampling layers, and adding a global residual connection. FERM has less inference time and more stable output. In order to process the various frequency components in the moire patterns, they propose a multi-scale feature encoding module (MSFE) that processes images at different scales. The MSFE has 3 simple versions of FERM with up and down sampling layers for different scales. In addition, they also propose an enhanced asymmetric convolution block (EACB), which can be applied to any image restoration network, see Figure 7. Based on ACB [14], they add two additional diagonal convolutions to further strengthen the kernel skeleton, and remove the batch normalization layers and the bias parameters for better performance. They used the proposed FERM and EACB to achieve the second place in the track 2: sRGB of the real image denoising challenge [1].

For the burst task, they propose a multi-frame spatial transformer network (M-STN). In order to connect multiple input frames, they concatenate n input frames together over the channel dimension (n is the number of input frames, including 1 standard frame and $n-1$ non-aligned frames). The
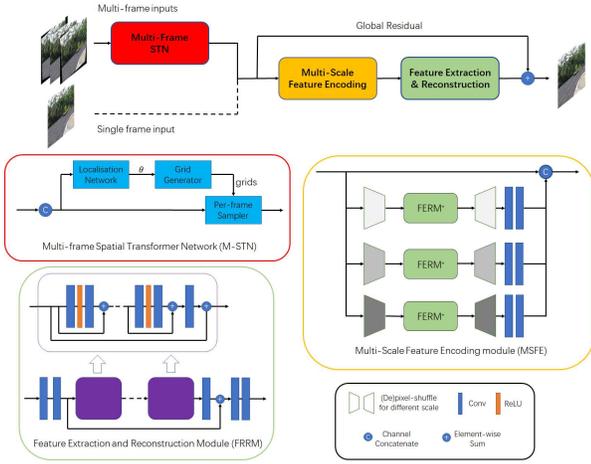
Figure 6: Network architecture of multi-frame and multi-scale for image demoireing (MMDM).
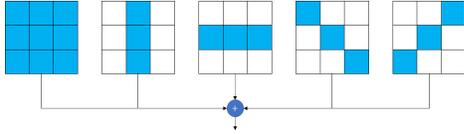
Figure 7: Enhanced asymmetric convolution block (EACB).

Figure 8: Overall structure of C3Net.
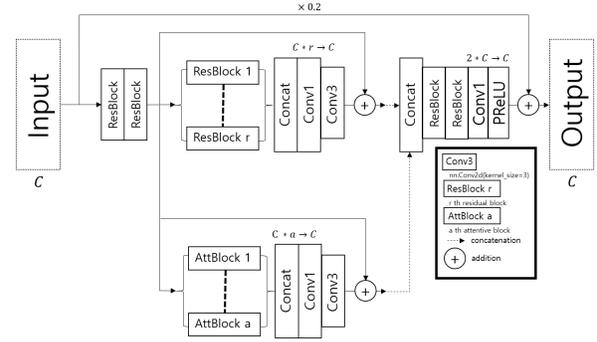
Figure 9: The structure of Attention-Via-Concat Block (AVCBlock).

localisation network processes the input to get $8 \times (n-1)$ parameters (they use the perspective transform which can simulate the burst input, and only transform the non-aligned frames), and constructs $n-1$ perspective transformation matrices. The grid generator and sampler respectively transform the input frames according to the matrices to obtain n aligned frames. The aligned frames are concatenated together over the channel dimension, then input to the main network. With more aligned high-frequency information, the performance of the network has been greatly improved. Compared to STN [24], the M-STN can process multi-frame inputs at the same time and make information fusion between them.

For both tracks, the team only used the simplified version of EACB (without diagonal convolutions) because EACB had a long training time. The use of EACB in this model, although it will consume more time during the training phase, the asymmetric convolutions can be fused, and there is no additional time consumption during testing phase.

Ensembles and fusion strategies: Both tracks use self-ensemble X8 (flip and rotate). After training, the weight of the asymmetric convolution is added to the corresponding position to achieve the fusion effect. This is not like a general fusion strategy, but more like self-fusion.

4.4. Reboot team

The **Reboot** team proposed *C3Net: Demoireing Network Attentive in Channel, Color and Concatenation* [27]. The method is inspired by Residual Non-local Attention Net-

work (RNAN) [63] and Deep Iterative Down-Up CNN for Image Denoising (DIDN) [57]. The entire network consists of n Attention Via Concatenation Blocks (AVCBlocks) in global feature fusion used in Residual Dense Network for Image Super-Resolution (RDN)[64], see Figures 8. An AVCBlock, see Figure 9, consists of two branches: the trunk branch and the mask branch. In the trunk branch, there are r residual blocks (ResBlocks) in parallel to retain the original values of input in diverse ways. In the mask branch, there are a attentive blocks (AttBlock) for guiding the values from trunk branch to demoire. Outputs from the two branches are concatenated and the number of channels of concatenated features is halved for entering next block. The method uses ResBlocks, each ResBlock similar to the one used in [56] which consists of one convolutional layer, one PReLU layer, and another convolutional layer. Channel attention [62] is added with ReLU to cope with demoireing problems related to colors. An AttBlock is a U-net with Resblocks, convolutional layers with stride 2 for downscaling, and pixel shuffle with kernel size 2 for upscaling features. The scaling layers and usage of U-net as a block benchmarked DIDN.

For burst processing, global maxpooling [3] is added following each AVCBlock. The algorithm gives feature maps which have maximum among 7 images and replicate and concatenate 7 times to match the dimension. The proposed C3Net concatenates the output of AVCBlock and the output of AVCBlock and global maxpooling layer and the number of channels of concatenated features halves for entering next block.

4.5. LIPT
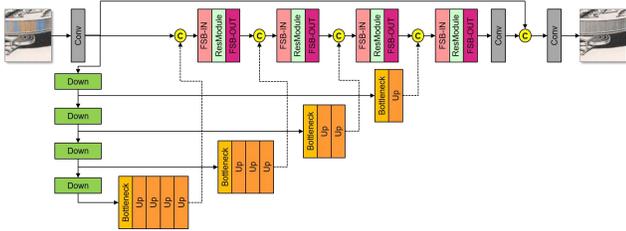
Figure 10: Overall architecture of the multi-scale deep residual network with adaptive pointwise convolution model.

(a) *FSB_IN*   (b) *FSB_OUT*   (c) *APC*
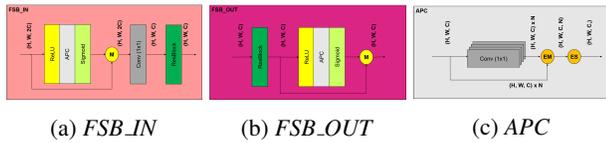
Figure 11: Proposed modules in the model.

The **LIPT** team proposed *Multi-scale deep residual network with adaptive pointwise convolution*. In order to extract more useful features, a novel adaptive pointwise convolution (APC) is proposed and applied, which is a modified version of a pointwise convolution based on 1×1 convolution filters [45]. The APC uses spatially variant learnable kernel weights for each pixel feature in order to perform adaptive linear projection. As shown in Figure 10, the overall architecture consists of two branches, namely, main and multi-scale branches. Features are extracted from the input image by applying a 3×3 convolution.

Each multi-scale branch performs down-scaling, feature extraction, and up-scaling processes by using *Down*, *Bottleneck*, and *Up* modules, respectively. It can handle multi-scale features efficiently. The *Down* module is a 3×3 convolution with a stride of 2. The *Up* module expands the feature map size by applying a 3×3 convolution and a sub-pixel convolution [41]. The *Bottleneck* module consists of a pointwise convolution with ReLU as the activation function, the *Up*, and the *Down* modules.

The extracted multi-scale information is merged into the main branch, which consists of several residual blocks (*ResBlocks*). Four-level multi-scale branches are used, thus, the *ResBlocks* are grouped into 4 *ResModules*. In order to extract important features from the main and multi-scale branches, feature selection blocks (*FSBs*) are used at the beginning and end of the each *ResModule*. The proposed *FSB_IN* and *FSB_OUT* are shown in Fig. 11. These include the *ResBlock* and the feature selection unit. The feature selection unit is modified from the selection unit [13]. The proposed *APC* is used instead of a 1×1 convolution in the selection unit. It can help to extract useful features from the

main and multi-scale branches along channels as a channel attention.

Adaptive pointwise convolution: Adaptive convolution has adaptive kernel weights that are learnable for each pixel feature [35, 42]. A new adaptive pointwise convolution (*APC*) is proposed, as shown in Figure 11c. The *APC* is a 1×1 convolution that has spatially variant learnable kernel weights for each pixel feature. Thus, the *APC* can obtain optimized feature maps by performing linear projection per pixel feature. The *APC* consists of N pointwise convolutions. In Figure 11c, *EM* denotes element-wise multiplication, also known as the Hadamard product between each output feature maps of the pointwise convolutions and input feature map. And *ES* denotes element-wise summation for the N feature maps. Note that the adaptive convolution kernel space is determined by output feature maps of N pointwise convolutions.

Three loss functions are used to train the model as $L = L_1 + L_{DCT} + L_\nabla$, where L_{DCT} is the DCT loss, L_∇ is the differential content loss. The L_{DCT} and L_∇ can help alleviate over-smoothness in the demoired image and tend to reconstruct image for high frequency components [12].

4.6. iipl-hdu team
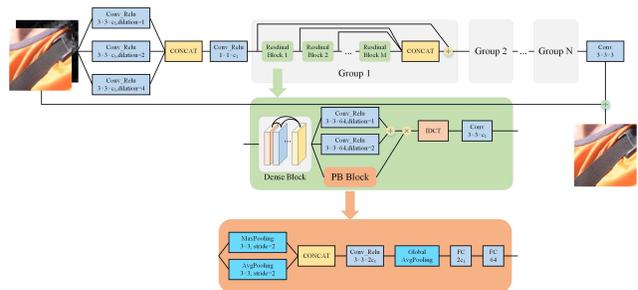
Figure 12: Architecture of the proposed baseline.
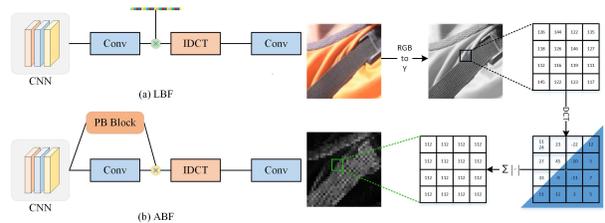
Figure 13: Proposed modules by **iipl-hdu** team. Left: the comparison of LBF (a) and ABF (b). Right: the pipeline of producing the label map.

The **iipl-hdu** team proposed a **CNN-based adaptive bandpass filter for image demoireing**, where Adaptive Bandpass Filte (ABF) is designed. This work is mainly inspired by the team iipl-hdu's previous work, the learnable bandpass filter [66, 65] (LBF). The new method intro-

duced a predictive bandpass (PB) block to adaptively predict the passband along with the image texture changing. Fig. 13 shows the difference between ABF and LBF. ABF was used to construct the residual block, and the architecture of GRDN [26] was adopted to construct the baseline. The architecture of proposed baseline is shown in Fig. 12. The baseline is constructed by N groups, and each group contains M residual blocks. In the residual block, let's define the input as a c_1-channel feature map, then the dense block includes 5 densely connected dilated convolution layers. The dilation rates of the 5 layers are $1, 2, 3, 2, 1$, and all of them output a c_2-channel feature map.

Label Map: Realizing that the moire effect only appears in high-frequency areas, a labeling method was proposed to generate the label map, highlighting the high-frequency area in the image. The label map will be concatenated to the moire image and sent to the CNN. The steps shown in Fig. 13 were followed to generate the label map from the moire input.

4.7. Mac AI team
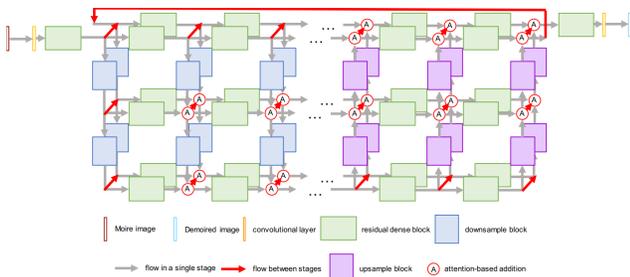
Figure 14: The main architecture of CubeDemoireNet.

The **Mac AI** team proposed *CubeDemoireNet: Enhanced Multi-Scale Network for Image Demoireing*. CubeDemoireNet is inspired by [30]. It enhances the conventional multi-scale network by 1) facilitating information exchange across different scales and stages to alleviate the underlying bottleneck issue, and 2) employs the attention mechanism to identify the dynamic Moiré patterns, and thus eliminate them effectively with the preservation of image texture.

Due to the fact that dynamic Moiré patterns usually embed in a broad range of frequencies, the benefit of a multi-scale structure is evident. However, the conventional multi-scale network has a hierarchical design that often suffers a bottleneck issue caused by insufficient information flow among different scales. the Mac AI team proposed a new network, named CubeDemoireNet, which promotes information exchange across different scales and stages via dense connections using upsampling/downsampling blocks to alleviate the underlying bottleneck issue prevalent in conventional multi-scale networks. Moreover, CubeDe-

moireNet further integrates the attention mechanism to preserve image texture while eliminating the dynamic Moiré patterns. Fig. 14 demonstrates the main architecture of the proposed CubeDemoireNet. The network consists of two convolutional layers and three blocks that are residual dense block (RDB), upsampling block, and downsampling block respectively. More specifically, the same settings in [64] are adopted in the proposed RDB. The upsampling and downsampling blocks are comprised of only two convolutional layers to adjust the spatial resolutions accordingly. As for the attention mechanism, the one employed in SENet [21] is chosen.

In single and burst demoireing tracks, the proposed CubeDemoireNet is used as the backbone. In CubeDemoireNet, the row, column, and stage is set to 3, 12, and 3 respectively for both tracks. Moreover, it was found that adding a refinement network can further improve the demoireing performance. Therefore, a modified U-net [38] is adopted for both tracks. Specifically, for the burst demoireing track, to align the burst images in accordance with the reference one, the Pyramid, Cascading and Deformable (PCD) proposed by [50] is employed.

4.8. MoePhoto team
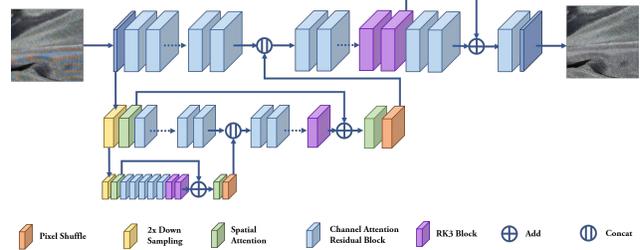
Figure 15: Overall structure of MSFAN.

The **MoePhoto** team proposed *Multi-scale feature aggregation network for image demoiréing*. This is a multi-scale feature aggregation network for image demoiréing, and dubbed MSFAN, see Figure 15. Based on [44, 11], MSFAN incorporates a multi-branch framework to encode the image with the original resolution features and $2\times$, $4\times$ down sampling features to obtain different feature representations. In each branch, local residual learning is used to remove moire pattern and recover image details. Then the features are upsampled with sub pixel convolution and concatenated to the upper level for effective feature aggregation. MSFAN uses channel attention and spatial attention to further enhance the information among the channels and grab the non-local spatial information among features with different spatial resolutions. L1 Charbonnier loss and L1 wavelet loss are used to calculate the difference between the output image and the ground truth and update the network parameters.
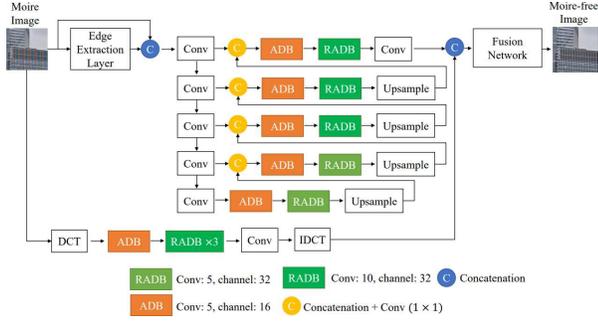
Figure 16: The architecture of the proposed dual-domain demoireing network for Track-1.

Figure 17: Proposed multilevel hyper vision net architecture.

4.9. DGU-CILAB

The **DGU-CILAB** team proposed *Dual-Domain Deep Convolutional Neural Networks for Demoireing* [49], which exploit the complex properties of moiré patterns in multiple domains, *i.e.*, pixel domain and frequency domain. In the pixel domain, multi-scale features are employed to remove the moiré artifacts associated with specific frequency bands through multi-resolution feature maps. In the frequency domain, inputs are transformed to spectral sub-bands using the discrete cosine transform (DCT). Then, DGU-CILAB designed a network that processes DCT coefficients to remove moiré artifacts. Next, a dynamic filter generation network [25] is developed to learn dynamic blending filters. Finally, the results from the pixel domain and frequency domain are combined by blending filters to yield the clean images. The **Pixel Network** is composed of multiple branches of different resolutions. Each branch is a stack of attention dense block (ADB) and residual attention dense block (RADB), which are based on convolutional block attention module (CBAM) [51], dense block (DB), and residual dense block (RDB) [64]. The **Frequency Network** processes the DCT coefficients of the input image to remove moiré artifacts in a frequency domain. The final output image is generated by applying the inverse discrete cosine transform (IDCT). The **Fusion Network** using dynamic filter network [25] is developed to take a pair of results from the pixel network and frequency network as input and outputs blending filters, which are then used to process the inputs to yield the final moiré-free image. To train the network, first the pixel network and the frequency network are trained separately. Then, after fixing them, the fusion network is trained. For the burst track, information of each image in the sequence is exploited and aligned with the reference image using the attention network.

4.10. Image Lab team

The **Image Lab** team proposed *Moire Image Restoration using Multilevel Hyper Vision Net*. The proposed architecture[39] is shown in Figure 17, where the input im-

age pixels mapped to the Cartesian coordinate space with the help of a coordinate convolution layer [28]. The output of coordinate convolution layers passed to encoder block. The convolutional layer and the residual dense attention blocks are utilized for better performance. The proposed network has the properties of an encoder and decoder structure of a vanilla U-Net [40, 38]. During down-sampling three blocks have been used in the encoder phase. In each block, the first encoder block is a 3×3 convolutional layer, followed by two residual dense attention blocks added and at the next block convolution layer with stride 2 used for down sampling. In the decoder phase, the same blocks have been used except the down sampling layer, which replaced with a super pixel convolutional layer. Output features of the encoder are fed to the convolution block attention module [51] (CBAM) and in the skip connection upsampled features concatenated with encoder block and the output of the CBAM block. Inspired by the Hyper Vision Net [40] model, in this work the three hyper vision layers in the decoder part were introduced. The output of these hyper vision layers are fused and supervised to obtain the enhanced image. The loss function is $Loss = MSE + (1 - SSIM) + SOBEL_{loss}$.

4.11. VAADER team

The **VAADER** team's method is inspired by MBCNN of the AIM2019 challenge [59] and uses a CNN-based multiscale approach. A multi loss extracted from the flow at different scales is used to train the network. A *pre-training* is performed on a mix of AIM2019 challenge patches [58] and this year challenge samples. The final model is obtained through a second training stage using only the current challenge dataset. Indeed, the past year dataset has several differences including large luminance shifts between input and reference that do not appear in this year's dataset.

4.12. CET_CVLab team

The **CET_CVLab** team proposed *Single Image Demoireing With Enhanced Feature Supervised Deep Convolutional Neural Network*. The architecture follows a CNN based encoder-decoder structure. The base network was proposed for image deraining [52] and has two convolutional encoders and one decoder i.e, two inputs and one output. Encoder-I is a sequence of downsamplers that extract the features of the degraded image at different scales. Encoder-II, which has similar structure as encoder-I, extracts the features of the moire-free image at different scales during training. The decoder is a sequence of upsamplers that reconstruct the moire-free images from encoder-I's output. The features of moire interfered image can be brought closer to that of moire-free image by minimizing their distance of separation in feature space. This is realized by incorporating feature loss during training. Feature loss is the mean of weighted sum of L1 distance between the features of the degraded image and moire-free image at different scales. To enhance the network, a block of residual dense network (RDN) [64] with channel attention (CA) [62] was introduced before downsampling and upsampling to extract the features in the current scale level. The proposed solution has 6 different scale levels. The model complexity can be adjusted by adding or removing the scale levels

4.13. sinashish team

The **sinashish** team proposed the use of feature fusion attention for the demoireing of moired images inspired by FFA-Net [36]. The method was trained with an $L2$ loss instead of the $L1$ loss proposed in the FFA-Net paper.

4.14. NTIREXZ team

The **NTIREXZ** team followed the work of FFA-Net [36] to estimate the undesired texture using global residual learning. Attention mechanisms, including both channel attention and pixel-wise attention, are used alongside a multi-scale approach for efficient learning.

5. Conclusion

This paper reviews the image demoireing challenge that was part of the NTIRE2020 workshop. First, the paper describes the challenge including the new dataset, CFAMoire that was created for participants to develop their demoire methods as well as evaluate proposed solutions. The challenge consisted of two tracks (single and burst) that considered different inputs to the demoire approaches submitted by participants. The tracks had 142 and 99 registered participants, respectively, and a total of 14 teams competed in the final testing phase. The paper summarizes the results of the two tracks, and then describes each of the approaches that

competed in the challenge. The entries span the current the state-of-the-art in the image demoireing problem. We hope this challenge and its results will inspire additional work in the demoire problem, which is becoming an increasingly important image quality challenge.

Acknowledgements

A. Teams

0. NTIRE 2020 team

- **Members:**
  Shanxin Yuan[1] (shanxin.yuan@huawei.com)
  Radu Timofte[2] (radu.timofte@vision.ee.ethz.ch)
  Aleš Leonardis[1] (ales.leonardis@huawei.com)
  Gregory Slabaugh[1] (gregory.slabaugh@huawei.com)
- **Affiliations:**
  [1] Huawei Noah's Ark Lab
  [2] ETH Zürich, Switzerland

1. HaiYun team

- **Members:** Xiaotong Luo (xiaotluo@qq.com)[1], Jiangtao Zhang[1], Ming Hong[1], Yanyun Qu[1], Yuan Xie[2], Cuihua Li[1]
- **Affiliations:**
  [1] Xiamen University, China
  [2] East China Normal University, China

2. OIerM team

- **Members:** Dejia Xu (dejia@pku.edu.cn)[1], Yihao Chu[2], Qingyan Sun[3]
- **Affiliations:**
  [1] Peking University, China
  [2] Beijing University of Posts and Telecommunications, China
  [3] Beijing Jiaotong University, China

3. Alpha team

- **Members:** Shuai Liu[1] (18601200232@163.com), Ziyao Zong[1], Nan Nan[1], Chenghua Li[2]
- **Affiliations:**
  [1] North China University of Technology, China
  [2] Institute of Automation, Chinese Academy of Sciences, China

4. Reboot team

- **Members:** Sangmin Kim (ksmh1652@gmail.com), Hyungjoon Nam, Jisu Kim, Jechang Jeong
- **Affiliation:** Hanyang University, Korea

5. LIPT team

- **Members:** Manri Cheon (manri.cheon@lge.com), Sung-Jun Yoon, Byungyeon Kang, Junwoo Lee
- **Affiliation:** LG Electronics

6. iipl-hdu team

- **Members:** Bolun Zheng (zhengbolun1024@163.com)
- **Affiliation:** Hangzhou Dianzi University, China

7. Mac AI team

- **Members:** Xiaohong Liu (liux173@mcmaster.ca), Linhui Dai, Jun Chen
- **Affiliation:** McMaster University, Canada

8. MoePhoto team

- **Members:** Xi Cheng (chengx@njust.edu.cn), Zhenyong Fu, Jian Yang

- **Affiliation:** Nanjing University of Science and Technology, China

9. DGU-CILAB team

- **Members:** Chul Lee (chullee@dongguk.edu), An Gia Vien, Hyunkook Park
- **Affiliation:** Dongguk University, South Korea

10. Image Lab team

- **Members:** Sabari Nathan (sabarinathantce@gmail.com)[1] , M.Parisa Beham[2] , S Mohamed Mansoor Roomi[3]
- **Affiliations:**
  [1] Couger Inc ,Tokyo,Japan
  [2] Sethu Institute of Technology, India
  [3] Thiagarajar college of Engineering, India

11. VAADER team

- **Members:** Florian Lemarchand (florian.lemarchand@insa-rennes.fr)[1], Maxime Pelcat[1], Erwan Nogues[1,2]
- **Affiliation:**
  [1] Univ. Rennes, INSA Rennes, IETR - UMR CNRS 6164, France
  [2] DGA-MI, Bruz, France

12. CET_CVLab team

- **Members:** Densen Puthussery (puthusserydensen@gmail.com), Hrishikesh P S, Jiji C V
- **Affiliation:** College of Engineering Trivandrum

13. Sinashish team

- **Members:** Ashish Sinha (asinha@mt.iitr.ac.in)
- **Affiliation:** Indian Institute of Technology Roorkee, India

14. NITREXZ team

- **Members:** Xuan Zhao (zxstud@163.com)
- **Affiliation:** Nanjing University of Aeronautics and Astronautics, China

References

[1] Abdelrahman Abdelhamed, Mahmoud Afifi, Radu Timofte, Michael Brown, et al. Ntire 2020 challenge on real image denoising: Dataset, methods and results. In *CVPRW, 2020.*

[2] Abdelrahman Abdelhamed, Radu Timofte, and Michael S. Brown. Ntire 2019 challenge on real image denoising: Methods and results. In *CVPRW, 2019.*

[3] Miika Aittala and Frédo Durand. Burst image deblurring using permutation invariant convolutional neural networks. In *ECCV, 2018.*

[4] Cosmin Ancuti, Codruta O Ancuti, and Radu Timofte. Ntire 2018 challenge on image dehazing: Methods and results. In *CVPRW, 2018.*

[5] Codruta O. Ancuti, Cosmin Ancuti, Florin-Alexandru Vasluianu, Radu Timofte, et al. Ntire 2020 challenge on non-homogeneous dehazing. In *CVPRW, 2020.*

[6] Boaz Arad, Radu Timofte, Yi-Tun Lin, Graham Finlayson, Ohad Ben-Shahar, et al. Ntire 2020 challenge on spectral reconstruction from an rgb image. In *CVPRW, 2020.*

[7] Jonathan Barron and Yun-Ta Tsia. Fast fourier color constancy. In *CVPR, 2017.*

[8] Yochai Blau, Roey Mechrez, Radu Timofte, Tomer Michaeli, and Lihi Zelnik-Manor. The 2018 pirm challenge on perceptual image super-resolution. In *ECCV, 2018.*

[9] Jianrui Cai, Shuhang Gu, Radu Timofte, and Lei Zhang. Ntire 2019 challenge on real image super-resolution: Methods and results. In *CVPRW, 2019.*

[10] Yue Cao, Jiarui Xu, Stephen Lin, Fangyun Wei, and Han Hu. Gcnet: Non-local networks meet squeeze-excitation networks and beyond. *arXiv, 2019.*

[11] Xi Cheng, Zhenyong Fu, and Jian Yang. Multi-scale dynamic feature encoding network for image demoiréing. *arXiv, 2019.*

[12] Manri Cheon, Jun-Hyuk Kim, Jun-Ho Choi, and Jong-Seok Lee. Generative adversarial network-based image super-resolution using perceptual content losses. In *ECCV, 2018.*

[13] Jae-Seok Choi and Munchurl Kim. A deep convolutional neural network with selection units for super-resolution. In *CVPRW, 2017.*

[14] Xiaohan Ding, Yuchen Guo, Guiguang Ding, and Jungong Han. Acnet: Strengthening the kernel skeletons for powerful cnn via asymmetric convolution blocks. In *ICCV, 2019.*

[15] Dario Fuoli, Zhiwu Huang, Martin Danelljan, Radu Timofte, et al. Ntire 2020 challenge on video quality mapping: Methods and results. In *CVPRW, 2020.*

[16] Michaël Gharbi, Gaurav Chaurasia, Sylvain Paris, and Frédo Durand. Deep joint demosaicking and denoising. *TOG, 2016.*

[17] Michaël Gharbi, Jiawen Chen, Jonathan T Barron, Samuel W Hasinoff, and Frédo Durand. Deep bilateral learning for real-time image enhancement. *TOG, 2017.*

[18] Shuhang Gu, Andreas Lugmayr, Martin Danelljan, Manuel Fritsche, Julien Lamour, and Radu Timofte. Div8k: Diverse 8k resolution image dataset. In *ICCVW, 2019.*

[19] Shuhang Gu and Radu Timofte. A brief review of image denoising algorithms and beyond. In *CiML, 2019.*

[20] Samuel W Hasinoff, Dillon Sharlet, Ryan Geiss, Andrew Adams, Jonathan T Barron, Florian Kainz, Jiawen Chen, and Marc Levoy. Burst photography for high dynamic range and low-light imaging on mobile cameras. *TOG, 2016.*

[21] Jie Hu, Li Shen, and Gang Sun. Squeeze-and-excitation networks. In *CVPR, 2018.*

[22] Andrey Ignatov, Nikolay Kobyshev, Radu Timofte, Kenneth Vanhoey, and Luc Van Gool. Dslr-quality photos on mobile devices with deep convolutional networks. In *ICCV, 2017.*

[23] Takashi Isobe1, Songjiang Li, Xu Jia, Shanxin Yuan, Ya-Li Li, Shengjin Wang, and Qi Tian. Video super-resolution with temporal group attention. In *CVPR, 2020.*

[24] Max Jaderberg, Karen Simonyan, Andrew Zisserman, and koray kavukcuoglu. Spatial transformer networks. In *NeurIPS, 2015.*

[25] Xu Jia, Bert De Brabandere, Tinne Tuytelaars, and Luc V Gool. Dynamic filter networks. In *NeurIPS, 2016.*

[26] Dong-Wook Kim, Jae Ryun Chung, and Seung-Won Jung. Grdn: Grouped residual dense network for real image denoising and gan-based real-world noise modeling. In *CVPRW, 2019.*

[27] Sangmin Kim, Hyungjoon Nam, Jisu Kim, and Jechang Jeong. C3net: Demoiréing network attentive in channel, color and concatenation. In *CVPRW, 2020.*

[28] Rosanne Liu, Joel Lehman, Piero Molino, Felipe Petroski Such, Eric Frank, Alex Sergeev, and Jason Yosinski. An intriguing failing of convolutional neural networks and the coordconv solution. In *NeurIPS, 2018.*

[29] Shuai Liu, Chenghua Li, Nan Nan, Ziyao Zong, and Ruixia Song. MMDM: Multi-frame and multi-scale for image demoiréing. In *CVPRW, 2020.*

[30] Xiaohong Liu, Yongrui Ma, Zhihao Shi, and Jun Chen. Grid-dehazenet: Attention-based multi-scale network for image dehazing. In *ICCV, 2019.*

[31] Andreas Lugmayr, Martin Danelljan, Radu Timofte, et al. Aim 2019 challenge on real-world image super-resolution: Methods and results. In *ICCVW, 2019.*

[32] Andreas Lugmayr, Martin Danelljan, Radu Timofte, et al. Ntire 2020 challenge on real-world image super-resolution: Methods and results. In *CVPRW, 2020.*

[33] Seungjun Nah, Sungyong Baik, Seokil Hong, Gyeongsik Moon, Sanghyun Son, Radu Timofte, and Kyoung Mu Lee. Ntire 2019 challenge on video deblurring and super-resolution: Dataset and study. In *CVPRW, 2019.*

[34] Seungjun Nah, Sanghyun Son, Radu Timofte, Kyoung Mu Lee, et al. Ntire 2020 challenge on image and video deblurring. In *CVPRW, 2020.*

[35] Simon Niklaus, Long Mai, and Feng Liu. Video frame interpolation via adaptive separable convolution. In *ICCV, 2017.*

[36] Xu Qin, Zhilin Wang, Yuanchao Bai, Xiaodong Xie, and Huizhu Jia. Ffa-net: Feature fusion attention network for single image dehazing. *arXiv, 2019.*

[37] Yaniv Romano, John Isidoro, and Peyman Milanfar. Raisr: rapid and accurate image super resolution. *TCI, 2016.*

[38] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *MICCAI, 2015.*

[39] D Sabarinathan, M Parisa Beham, and SM Roomi. Moire image restoration using multi level hyper vision net.

[40] D Sabarinathan, M Parisa Beham, SM Roomi, et al. Hyper vision net: Kidney tumor segmentation using coordinate convolutional layer and attention unit. *arXiv, 2019.*

[41] Wenzhe Shi, Jose Caballero, Ferenc Huszár, Johannes Totz, Andrew P Aitken, Rob Bishop, Daniel Rueckert, and Zehan Wang. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In *CVPR, 2016.*

[42] Hang Su, Varun Jampani, Deqing Sun, Orazio Gallo, Erik Learned-Miller, and Jan Kautz. Pixel-adaptive convolutional neural networks. In *CVPR, 2019.*

[43] Baochen Sun and Kate Saenko. Deep coral: Correlation alignment for deep domain adaptation. In *ECCV, 2016.*

[44] Yujing Sun, Yizhou Yu, and Wenping Wang. Moiré photo restoration using multiresolution convolutional neural networks. *TIP, 2018.*

[45] Zhi Tian, Tong He, Chunhua Shen, and Youliang Yan. Decoders matter for semantic segmentation: Data-dependent decoding enables flexible feature aggregation. In *CVPR, 2019.*

[46] Radu Timofte, Eirikur Agustsson, Luc Van Gool, Ming-Hsuan Yang, and Lei Zhang. Ntire 2017 challenge on single image super-resolution: Methods and results. In *CVPRW, 2017.*

[47] Radu Timofte, Shuhang Gu, Jiqing Wu, Luc Van Gool, Lei Zhang, Ming-Hsuan Yang, Muhammad Haris, et al. Ntire 2018 challenge on single image super-resolution: Methods and results. In *CVPRW, 2018.*

[48] Radu Timofte, Rasmus Rothe, and Luc Van Gool. Seven ways to improve example-based single image super resolution. In *CVPR, 2016.*

[49] An Gia Vien, Hyunkook Park, and Chul Lee. Dual-domain deep convolutional neural networks for image demoireing. In *CVPRW, 2020.*

[50] Xintao Wang, Kelvin CK Chan, Ke Yu, Chao Dong, and Chen Change Loy. Edvr: Video restoration with enhanced deformable convolutional networks. In *CVPRW, 2019.*

[51] Sanghyun Woo, Jongchan Park, Joon-Young Lee, and In So Kweon. Cbam: Convolutional block attention module. In *ECCV, 2018.*

[52] Peng Xiang, Lei Wang, Fuxiang Wu, Jun Cheng, and Mengchu Zhou. Single-image de-raining with feature-supervised generative adversarial network. *SPL, 2019.*

[53] Luo Xiaotong, Zhang Jiangtao, Hong Ming, Qu Yanyun, Xie Yuan, and Li Cuihua. Deep wavelet network with domain adaptation for single image demoireing. In *CVPRW, 2020.*

[54] Dejia Xu, Yihao Chu, and Qingyan Sun. Moiré pattern removal via attentive fractal network. In *CVPRW, 2020.*

[55] Wenhan Yang, Shiqi Wang, Dejia Xu, Xiaodong Wang, and Jiaying Liu. Towards scale-free rain streak removal via self-supervised fractal band learning. In *AAAI, 2020.*

[56] Songhyun Yu and Jechang Jeong. Local excitation network for restoring a jpeg-compressed image. *IEEE Access, 2019.*

[57] Songhyun Yu, Bumjun Park, and Jechang Jeong. Deep iterative down-up cnn for image denoising. In *CVPRW, 2019.*

[58] Shanxin Yuan, Radu Timofte, Gregory Slabaugh, and Ales Leonardis. Aim 2019 challenge on image demoireing: dataset and study. In *ICCVW, 2019.*

[59] Shanxin Yuan, Radu Timofte, Gregory Slabaugh, Ales Leonardis, et al. Aim 2019 challenge on image demoireing: methods and results. In *ICCVW, 2019.*

[60] Kai Zhang, Shuhang Gu, Radu Timofte, et al. Ntire 2020 challenge on perceptual extreme super-resolution: Methods and results. In *CVPRW, 2020.*

[61] Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, and Lei Zhang. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *TIP, 2017.*

[62] Yulun Zhang, Kunpeng Li, Kai Li, Lichen Wang, Bineng Zhong, and Yun Fu. Image super-resolution using very deep residual channel attention networks. In *ECCV, 2018.*

[63] Yulun Zhang, Kunpeng Li, Kai Li, Bineng Zhong, and Yun Fu. Residual non-local attention networks for image restoration. *arXiv, 2019.*

[64] Yulun Zhang, Yapeng Tian, Yu Kong, Bineng Zhong, and Yun Fu. Residual dense network for image super-resolution. In *CVPR, 2018.*

[65] Bolun Zheng, Yaowu Chen, Xiang Tian, Fan Zhou, and Xuesong Liu. Implicit dual-domain convolutional network for robust color image compression artifact reduction. *TCSVT, 2019.*

[66] Bolun Zheng, Shanxin Yuan, Gregory Slabaugh, and Aleš Leonardis. Image demoireing with learnable bandpass filters. In *CVPR, 2020.*