MMDM: Multi-frame and Multi-scale for Image Demoiréing

Shuai Liu[1], Chenghua Li[2,*], Nan Nan[1], Ziyao Zong[1], Ruixia Song[1]
[1]North China University of Technology, [2]Institute of Automation, Chinese Academy of Sciences
18601200232@163.com, lichenghua2014@ia.ac.cn, 2829272117@qq.com
824924664@qq.com, songrx@ncut.edu.cn

Abstract

The imaging characteristics of digital sensors often lead to the moiré patterns, which are widely distributed over the frequency domain and have irregular colors and shapes. The images with moiré patterns could lead to a serious decline in the visual quality. The difficulty of demoiréing lies in that the moiré patterns mix both low and high frequency information to be processed. In this paper, we propose MMDM, an effective image demoiréing network, which uses multiple images as inputs and multi-scale feature encoding module as low-frequency information enhancement. Our MMDM has three key modules: the newly designed multi-frame spatial transformer networks (M-STN), the multi-scale feature encoding module (MSFE), and the enhanced asymmetric convolution block (EACB). Especially, the M-STN aims to align the multiple input images simultaneously. The MSFE is for multiple frequency information encoding, which is built on the efficient EACB module. Experiments prove the effectiveness of MMDM. Also, our model achieves the 2nd place on both demoiring track and denoising track in the NTIRE2020 Challenge. Code is avaliable at: https://github.com/q935970314/MMDM

1. Introduction

Moiré is an irregular colored stripe that often appears on devices such as digital cameras or scanners. Mathematically speaking, two equal-amplitude sine waves with close frequencies are superimposed, and the amplitude of the synthesized signal will change according to the difference between the two frequencies. When the spatial frequency between pixels of the photosensitive element and the stripes in the image is close, moiré will occur. Moiré is irregular, has no obvious shape rule, and the frequency domain segment is wider. Compared to other image restoration tasks such as super-resolution and denoising, demoiréing is more difficult, and needs to process both low and high frequency patterns.

With the development of deep learning, convolutional neural network (CNN) has shown superior performance in computer vision tasks such as image classification [6], object detection [16] and natural language processing [3]. In image restoration tasks such as color enhancement, denoising and super-resolution, CNN-based methods can better fit the complex mapping between input image and ground truth, and restore the low and high frequency information of the image.

In this paper, we propose an effective network for image demoiréing, named as MMDM, which uses multi-frame as input and multi-scale feature encoding module as low-frequency information enhancement. The multi-frame images are captured in burst mode. We develop an improved STN module (named M-STN) to process the set of images. The M-STN can align the multi-frame inputs to a selected standard image simultaneously, by performing perspective transformations inspectively.

For feature encoding, we design an enhanced asymmetric convolution block, named EACB. Our EACB differs from the original asymmetric convolution block in adding two additional diagonal convolutions and removing the batch normalization layers, which makes it work better for image restoration tasks. Based on EACB, we build the final feature extraction and reconstruction module (FERM), and a simple version (FERM*) for multiple scales feature encoding module (MSFE). The outputs of the multiple scale features are catenated and then processed by FERM. The pipeline of our model is shown in Figure 1.

Furthermore, our submodules such as FERM, M-STN, and EACB would have better performances on other image restoration tasks. Compared with other methods, the proposed MMDM has better performance on objective indicator and visual effects. In the NTIRE2020 Demoiréing challenge, the proposed MMDM achieves the 3rd place in Track 1: Single image and the 2nd place in Track 2: Burst. And in the Denoising challenge, the proposed FERM with simple version of EACB achieves the 2nd place for Track 2: sRGB. In summary, the main contributions of this paper

*Corresponding author.

are as follows:

- We propose a new demoiréing network named MMDM which uses multiple images as inputs and multiple scale enhancement.
- We propose a M-STN which can perform perspective transformation alignment on multiple frames of input simultaneously.
- We propose a MSFE which can enhance the low-frequency information at different scales.
- We propose a EACB to improve the performance of image restoration networks.
- In the NTIRE2020 challenge, the proposed methods achieve 2nd place on the burst track of demoiréing [22], 2nd place on the sRGB track of denoising [1], and 3rd place on the single track of demoiréing.

2. Related work

Image restoration is one of the most important tasks in computer vision, and it is divided into many sub-tasks, such as denoising [9] and super-resolution [10, 26] which mainly focus on high-frequency details, and color enhancement which mainly focuses on low-frequency patterns. According to the characteristics of moiré generation, the task of demoiréing requires to pay attention to both low and high frequency patterns [2], so demoiréing is more difficult when compared with other tasks.

Demoiréing network. According to the multi-frequency characteristic of moiré, Sun et al. [19] propose a multi-resolution fully convolutional network (DMCNN) which processes different resolution images on different branches and adds the obtained results. Based on U-Net, Gao et al. [5] propose a multi-scale feature enhancing network (MSFE) which can process information at different scales, and add feature enhancement modules at each scale to fuse low-scale information into high-scale. Cheng et al. [2] propose a multi-scale dynamic feature encoding network (MDDM) which also uses a multi-branch structure to process information at different scales, and propose a Dynamic Feature Encoding module to deal with the dynamically changing moiré patterns.

High-frequency detail processing network. Zhang et al. [26] propose a residual channel attention network (RCAN) for the super-resolution task. In order to train very deep networks, they propose a residual group (RG) and a long/short skip connection. In order to make the network pay more attention to the correlation between channels, they propose a channel attention (CA) module. RCAN has a good effect in super-resolution task, it can process the high-frequency detail better and provide more details for the final reconstruction.

Spatial Transformer Networks. Max et al. [8] propose a spatial transformer network (STN). The transformation parameters generated by the local network are used by the grid generator to generate the sampling grid, and finally the sampler transforms the input according to the grid. STN enables the network to perform spatial transformations on inputs such as perspective transformations, rotation transformations, etc. It is important that STN is a plug-and-play module which can achieve the required results without any additional supervision information. STN has been widely used in face recognition and other fields.

Asymmetric Convolution. Ding et al. [4] propose an asymmetric convolution block (ACB) which enhances the CNN's kernel skeleton by adding additional 1×3 and 3×1 convolutions during training, and fuse the three convolutions during inference phase without any additional inference time. It has a good performance improvement in the field of classification and segmentation.

3. Proposed method

We propose multi-frame and multi-scale for image demoiréing (MMDM), an effective image demoiréing network, which uses multi-frame as inputs and multi-scale feature encoding module as low-frequency information enhancement. The detail structures are shown in Figure 1.

3.1. Network architecture

MMDM has three key modules: the newly designed multi-frame spatial transform network (M-STN), the multi-scale feature encoding module (MSFE), and the enhanced asymmetric convolution block (EACB). Especially, the M-STN is designed to align the multiple input frames simultaneously. The MSFE is built for multiple frequency information encoding. The main feature extraction and reconstruction module (FERM) is built on the new EACB, which suits for image restoration tasks than the original ACB [4].

The pipeline of MMDM is as follows. Firstly, taking multiple frames (commonly using burst mode of a camera) as the inputs, concatenating them over the channel dimension, and feeding into the M-STN. The M-STN would output the aligned images, with same channels and spatial size. Secondly, the MSFE takes the aligned images and encodes the multi-scale feature (1/2, 1/4, 1/8 of the original spatial size). Then, the multi-scale encoding features are concatenated and processed by the final FERM. Finally, an element-wise summation of the outputs of FERM and M-STN lead to the output result. Additionally, our pipeline can also process single image input, with removing the M-STN. The details are as follows.

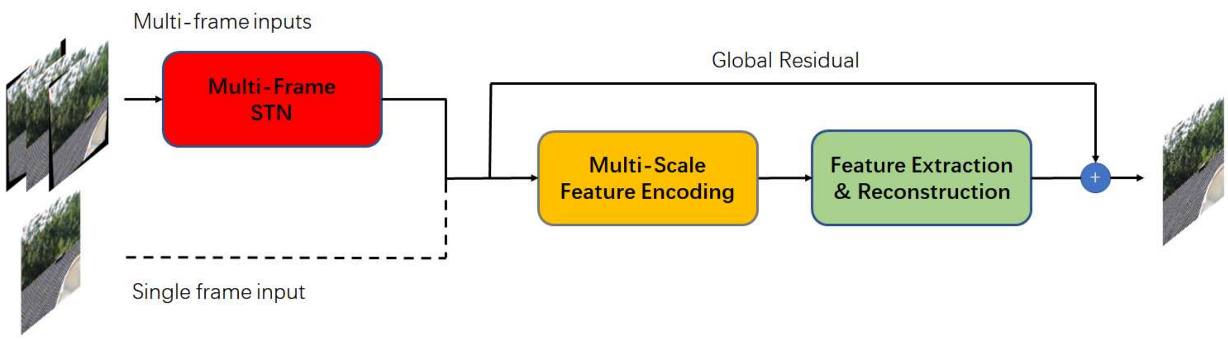


Figure 1. Network architecture of multi-frame and multi-scale for image demoiréing (MMDM)

3.1.1 Multi-frame Spatial Transformer Network

For the burst mode, the inputs are multiple frames at different angles of the same scene, including one standard frame and multiple non-aligned frames. Firstly, we align all the frames with the standard frame, and then perform demoiré processing. Perspective transformation can simulate the same scene at different angles, so we use the perspective transformation version of spatial transformer network (STN) [8] for alignment.

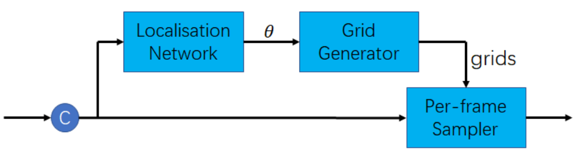


Figure 2. Network architecture of multi-Frame spatial transformer network (M-STN)

STN generally transforms a single frame of input adaptively, so we propose the multi-frame spatial transformer network (M-STN), as shown in Figure 2. In order to connect multiple input frames, we 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 localisation network processes the input to get $8 \times (n-1)$ parameters (we 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 modified images. Compared to the single frame alignment, multiple frame simultaneous alignment can increase the fusion of information between different frames. The proposed M-STN can be expressed as:

$$G_1, G_2, \cdots, G_{n-1} = GG(LN(cat(I_1, I_2, \cdots, I_n)))$$

$$Output_i = S(I_i, G_i) \quad i = 1, 2, \cdots, n-1$$

where I, cat, LN, GG, G, S denotes input frame, channel concatenate, localisation network, grid generator, grid and sampler respectively.

The aligned frames are concatenated together over the channel dimension, then as input to the main network. With more aligned high-frequency details, the performance of the network has been greatly improved. We visualize the effect of M-STN, as shown in Figure 3. Obviously, the transformed frames have been almost aligned with the standard frame over the spatial dimension.

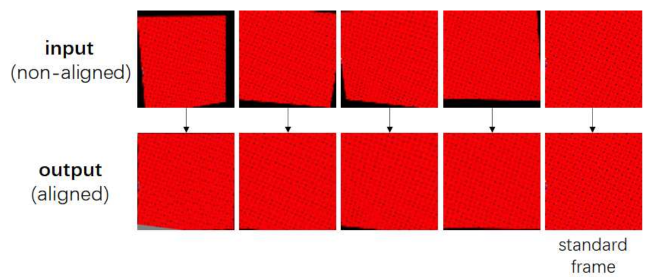


Figure 3. Visualization of input (non-aligned) and output (aligned) of multi-frame spatial transformer network (M-STN).

The proposed M-STN can be applied to any image restoration networks to align burst input.

3.1.2 Feature Extraction and Reconstruction Module

Because moiré's low-frequency patterns are relatively simple, we propose the feature extraction and reconstruction module (FERM) to mainly process the high-frequency details. The inspiration of FREM comes from RCAN [26] which is deep and suitable for processing high-frequency details. As shown in Figure 4, we improve as follows:

Remove all Channel Attention (CA) modules. The CA adaptively adjusts features through the interdependence between channels, which can enhance the performance of the network. However, we found that too many CAs will increase the training and inference time, but bring little performance improvement, as shown in Table 1. So we remove all CAs, after removing, the network is simpler and faster. The new residual block and residual group can be expressed as:

$$B(x) = Conv_1(ReLU(Conv_2(x))) + x$$

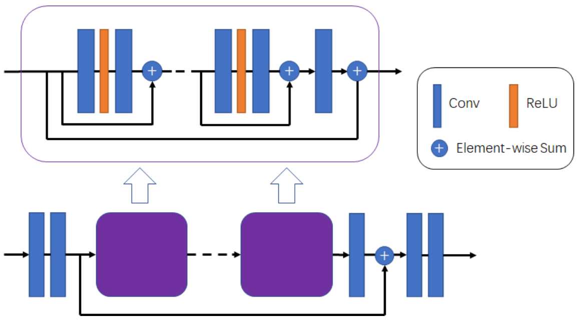


Figure 4. Network architecture of the feature extraction and reconstruction module (FERM)

Model	PSNR	Runtime (s)
RCAN	37.90	0.1124
RCAN w/o CA	37.87	0.0508

Table 1. Investigation of channel attention (CA) module, note that the PSNR results are from the paper of RCAN [26] and the input size is 128×128.

$$RG(x) = Conv(B_1(B_2(\cdots B_n(x)))) + x$$

where B, RG, $Conv$, $ReLU$ denotes residual block, residual group, convolution layer and activation function respectively.

Remove the upsampling module. Unlike the super-resolution task, the input and output of demoiréing task are the same size, so we remove the upsampling module to keep all of the features at the same size.

Add global residual. Global residual is increasingly used in image restoration networks [9, 21, 15], which can enhance the stability and make the network have higher performance in the early stages of training. So we add global residual to our network.

The proposed feature extraction and reconstruction module can be used as a backbone of any image restoration network, and has a good performance.

3.1.3 Multi-Scale Feature Encoding module

Due to the characteristics of moiré, the network must pay attention to the low frequency patterns [2]. As shown in Figure 5, the multi-scale feature encoding module (MSFE) has 3 simple versions of feature extraction and reconstruction module (FERM*) with up and down sampling layers. The MSFE down-samples the input at different scales firstly, then processes the corresponding low-frequency patterns, and finally performs up-sampling to resize the features to their original size for fusion. Pixelshuffle [18] is used for up-sampling and de-pixelshuffle [20] is used for down-sampling. Compared with deconvolution [24] and interpolation methods, pixelshuffle can preserve the information of original image better. We visualize the output features of MSFE, as shown in Figure 6. It is obvious that our proposed MSFE can extract the low-frequency patterns of moiré well.

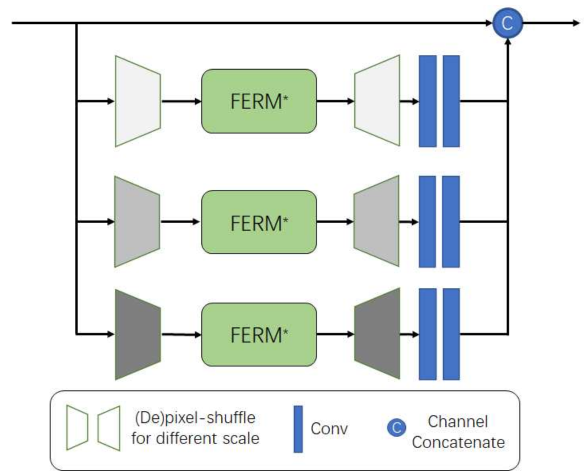


Figure 5. Network architecture of the multi-scale feature encoding module (MSFE), note that FERM* is a simple version of proposed feature extraction and reconstruction module with fewer parameters.

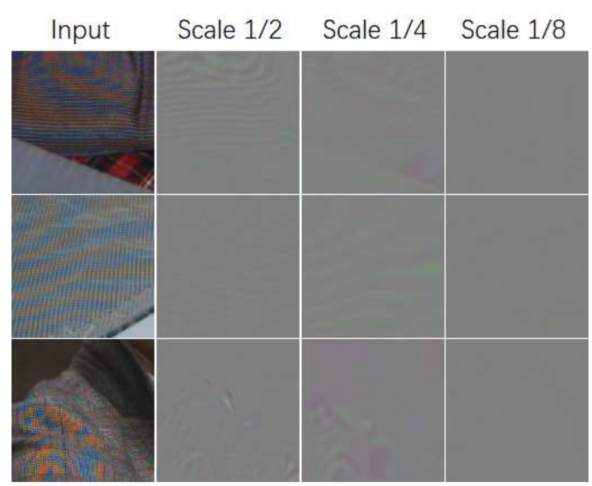
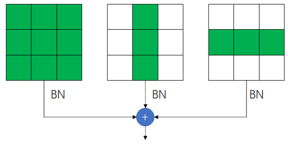


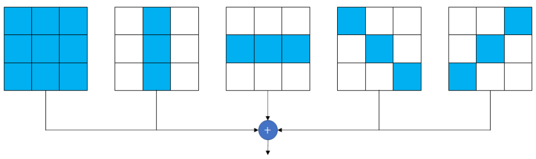
Figure 6. Visualization of the multi-scale feature encoding module (MSFE) output features at different scales.

3.2. Enhanced Asymmetric Convolution Block for image restoration

Asymmetric convolution block (ACB) [4] adds an additional 1×3 and an additional 3×1 convolution on the basis of square convolution to enhance the convolution kernel skeleton, batch normalization (BN) [7] is performed after each convolution, and the outputs are added as a new output. We express ACB as follow, where $+$ denotes element-wise



(a) ACB



(b) EACB

Figure 7. Comparison of (a) original asymmetric convolution block (ACB) and (b) proposed enhanced asymmetric convolution block (EACB). Note that EACB use the convolution without bias.

sum:

$$\begin{aligned} ACB(x) = BN_{3\times3}(Conv_{3\times3}(x)) + \\ BN_{1\times3}(Conv_{1\times3}(x)) + \\ BN_{3\times1}(Conv_{3\times1}(x)) \end{aligned}$$

Based on ACB, we add two additional diagonal convolutions (left diagonal and right diagonal) to further strengthen the kernel skeleton. The 1×3 and 3×1 convolutions in ACB can be considered as additional prior information, they focus on horizontal and vertical features better. We added two additional diagonal convolutions to focus on extracting oblique features.

BN is often used to improve the generalization ability of the network and speed up the training, allowing the network to be trained with a large learning rate. However, BN will reduce the specificity of image restoration networks [13], so we remove the BN layer in ACB. Without the correction of BN, the feature deviation is more serious and the network is easy to explode gradients during training, so we futher remove the bias parameters of all convolutional layers and use cosine annealing learning rate scheduler [14]. As shown in Figure 7, the new ACB for image restoration named as enhanced asymmetric convolution block (EACB), which can be expressed as:

$$EACB(x) = ACB^{*}(x) + Conv^{*}_{ldiag}(x) + Conv^{*}_{rdiag}(x)$$

The proposed EACB can be applied to any image restoration network to enhance the performance without any extra inference time.

4. Experiments

4.1. Dataset

We use the dataset provided by the NTIRE 2020 demoiréing challenge for experiments. It includes 10,000 pairs of training images, 500 pairs of validation images and 500 pairs of test images. The generation of data is similar to the LCDMoire dataset[23]. Since the ground truth of the testset is not released, we compare the methods on the validation set.

4.2. Training details and model settings

We use the Adam [11] optimizer with default parameters and the Charbonnier loss [12] to train our model. The initial learning rate is $1e-4$, and we use the cosine annealing learning rate scheduler [14] with about 300K iterations, the minimum learning rate is $5e-6$. The batch size is 16 and patch size is 128. We use PyTorch 1.3, NVIDIA RTX 2080Ti GPU with CUDA10.0 to accelerate training.

The proposed feature extraction and reconstruction module (FERM) contains 10 residual groups and each group contains 20 residual blocks (same as RCAN[26]). The proposed multi-scale feature encoding module (MSFE) contains 3 simplified version of FERM (name as FERM*) with defferent scales (1/2, 1/4, 1/8), each FERM* contains 4 residual groups and each group contains 8 residual blocks. We only use the simplified version of enhanced asymmetric convolution block (EACB without diagonal convolutions) to replace every convolutions in the model. There are 5 frames for multi-frame inputs (including 1 standard frame and 4 non-aligned frames).

4.3. Comparison with the state of the arts

To prove the effectiveness of our proposed model, we compared the model qualitatively and quantitatively with other deep learning-based methods. To be fair, all methods use the same training data and training parameters. The comparative models include DnCNN [25], VDSR [10], U-Net [17], DMCNN [19]. We use the average PSNR and SSIM of the validation set to compare objective performance. As shown in Table 2, our model is much higher than other models in both PSNR and SSIM. We also compared visual effects, as shown in Figure 8. Our model can reconstruct the details more accurately.

5. Ablation study

In this section, we perform ablation experiments on M-STN, MSFE and EACB to prove the effectiveness of the proposed module.

5.1. Multi-frame Spatial Transformer Network

The baseline is a small version of the feature extraction and reconstruction module with 5 residual groups. The input of the baseline is a single standard frame. The multiple frames are concatenated over the channel dimension. The curve of the validation PSNR is shown in Figure 9. It can be seen that concatenating the multiple frames as input directly does not bring much performance improvement. But

Method	Moiré	VDSR	U-Net	DMCNN	DnCNN	**MMDM** (single frame)	**MMDM** (multi-frame)
PSNR	25.1845	30.1090	35.2384	36.6919	37.4429	**41.9667**	**45.4625**
SSIM	0.7697	0.9529	0.9674	0.9774	0.9813	**0.9928**	**0.9968**

Table 2. Comparison of proposed MMDM and other demoiréing methods: DnCNN, VDSR, U-Net, and DMCNN.

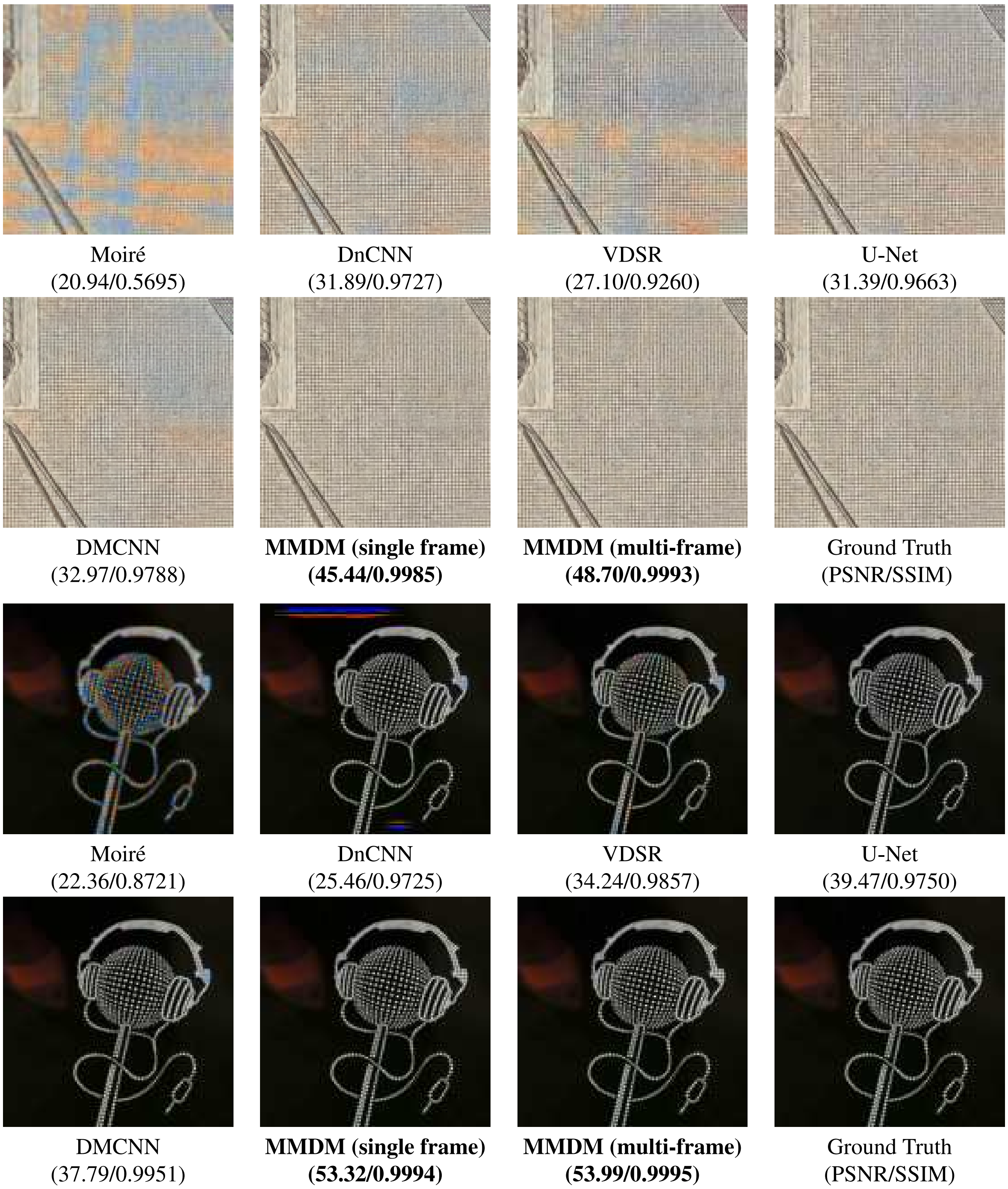


Figure 8. Comparison of proposed MMDM and other demoiréing methods: DnCNN, VDSR, U-Net, and DMCNN.

taking the aligned frames by the multi-frame spatial transformer network as input, the performance has a significant improvement.

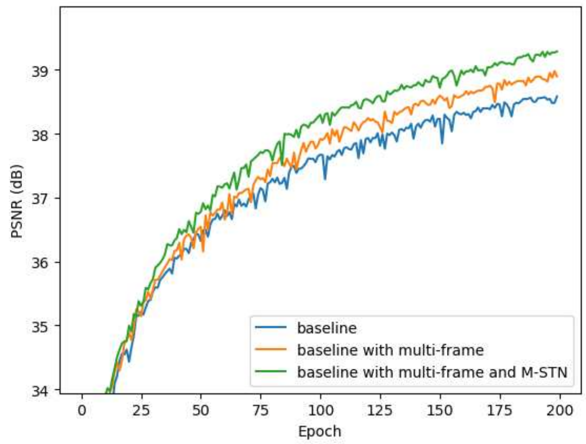


Figure 9. Effectiveness study of multi-frame spatial transformer networks (M-STN)

5.2. Multi-Scale Feature Encoding module

The FERM* in the multi-scale feature encoding module (MSFE) contains 2 residual groups. The curve of the validation PSNR is shown in Figure 10. When there is only a processing of scale 2, the performance improvement is not obvious. With the increasing of multi-scale, the performance is gradually improving. This proves the importance of multi-scale information and the effectiveness of proposed MSFE.

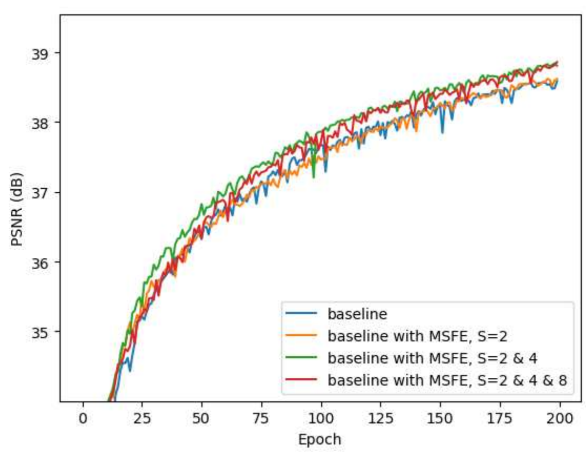


Figure 10. Effectiveness study of multi-scale feature encoding module (MSFE), note that S denotes scale

5.3. Enhanced Asymmetric Convolution Block

We use the enhanced asymmetric convolution block (EACB) to replace all convolutions of the baseline. The curve of the validation PSNR is shown in Figure 11. It can be seen that the baseline with the original ACB drops a lot of performances, but the baseline with EACB has a great performance improvement.

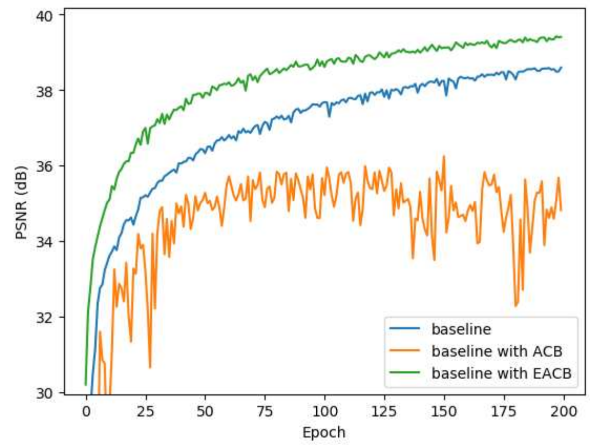


Figure 11. Effectiveness study of proposed enhanced asymmetric convolution block (EACB)

6. NTIRE 2020 challenges

New Trends in Image Restoration and Enhancement (NTIRE) in conjunction with CVPR 2020 has challenges on demoiréing [22] and denoising [1]. For the demoiréing challenge, the proposed MMDM with simple version of EACB achieves the 3rd place in Track 1: Single image and the 2nd place for Track 2: Burst.

It should note that due to time, our track 2 model is obtained by finetune training the track 1 model, which has only a small improvement. We retrained the track 2 model and got better performance (Figure 2). Because the single image model has been able to process the standard frame well, when finetune training the burst model, the M-STN has no effect and the model ignores the information of extra frames. The retraining has made M-STN effective, and the model performance has been greatly improved. We also test the retrained model on the test set, the results are shown in Table 3.

For the denoising challenge, we only use the proposed feature extraction and reconstruction module (FERM) with simple version of enhanced asymmetric convolution block, and achieve the 2nd place for Track 2: sRGB. The results are shown in Table 4. This also proves the effectiveness of proposed FERM and EACB.

Method	PSNR	SSIM	Method	PSNR	SSIM
1st	42.14	0.99	**MMDM***	45.32	1.00
2nd	41.95	0.99	1st	41.95	0.99
MMDM	41.84	0.99	**MMDM**	41.88	0.99
4th	41.11	0.99	3rd	40.64	0.99
5th	41.04	0.99	4th	40.33	0.99

Table 3. Results for NTIRE 2020 Demoiréing Challenge. Left: Track 1: Single image, right: Track 2: Burst. Note that MMDM* is the retrained model.

Method	PSNR	SSIM
1st	33.22	0.9596
FERM	33.12	0.9578
3rd	33.01	0.9590
4th	32.80	0.9565
5th	32.69	0.9572

Table 4. Results of NTIRE 2020 real image denoising challenge, Track 2: sRGB. Note that FERM is the proposed feature extraction and reconstruction module (FERM) with the simple version of enhanced asymmetric convolution block (EACB).

7. Conclusion

In this paper, we propose a image demoiréing network MMDM, which could take multiple frames as inputs and has multi-scale feature encoding ability. The proposed three key modules (M-STN, MSFE, and EACB) of MMDM is proved to be highly effective for image restoration tasks, such as demoiréing, denoise and so on. Our MMDM and simple versions achieve good results in both image demoiring and denoise. Later, we will further study the effect of EACB in lightweight image restoration tasks.

Acknowledgement

This work was supported in part by National Key Research and Development Program (2017YFF0209806); Advance Research Program (31511130301) and National Natural Science Foundation of China (No.61906193; No.61906195).

References

[1] Abdelrahman Abdelhamed, Mahmoud Afifi, Radu Timofte, Michael S. Brown, et al. Ntire 2020 challenge on real image denoising: Dataset, methods and results. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2020. 2, 7

[2] Xi Cheng, Zhenyong Fu, and Jian Yang. Multi-scale dynamic feature encoding network for image demoiréing. *The IEEE International Conference on Computer Vision (ICCV) Workshops*, 2019. 2, 4

[3] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018. 1

[4] Xiaohan Ding, Yuchen Guo, Guiguang Ding, and Jungong Han. Acnet: Strengthening the kernel skeletons for powerful cnn via asymmetric convolution blocks. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1911–1920, 2019. 2, 4

[5] Tianyu Gao, Yanqing Guo, Xin Zheng, Qianyu Wang, and Xiangyang Luo. Moiré pattern removal with multi-scale feature enhancing network. In *2019 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*, pages 240–245. IEEE, 2019. 2

[6] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Sun Jian. Deep residual learning for image recognition. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2016. 1

[7] Sergey Ioffe and Christian Szegedy. Batch normalization: accelerating deep network training by reducing internal covariate shift. 2015. 4

[8] Max Jaderberg, Karen Simonyan, Andrew Zisserman, et al. Spatial transformer networks. In *Advances in neural information processing systems*, pages 2017–2025, 2015. 2, 3

[9] 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. 2019. 2, 4

[10] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee. Accurate image super-resolution using very deep convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016. 2, 5

[11] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *Computer Science*, 2014. 5

[12] Wei Sheng Lai, Jia Bin Huang, Narendra Ahuja, and Ming Hsuan Yang. Fast and accurate image super-resolution with deep laplacian pyramid networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017. 5

[13] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. Enhanced deep residual networks for single image super-resolution. 5

[14] Ilya Loshchilov and Frank Hutter. SGDR: Stochastic gradient descent with warm restarts. 5

[15] Bumjun Park, Songhyun Yu, and Jechang Jeong. Densely connected hierarchical network for image denoising. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2019. 4

[16] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster R-CNN: Towards real-time object detection with region proposal networks. In *Advances in Neural Information Processing Systems (NIPS)*, 2015. 1

[17] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. 2015. 5

[18] 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. 2016. 4

[19] Y Sun, Y Yu, and Wp Wang. Moiré photo restoration using multiresolution convolutional neural networks. 2, 5

[20] Thang Vu, Cao V. Nguyen, Trung X. Pham, Tung M. Luu, and Chang D. Yoo. Fast and efficient image quality enhancement via desubpixel convolutional neural networks. 2018. 4

[21] Songhyun Yu, Bumjun Park, and Jechang Jeong. Deep iterative down-up cnn for image denoising. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2019. 4

[22] Shanxin Yuan, Radu Timofte, Ales Leonardis, Gregory Slabaugh, et al. Ntire 2020 challenge on image demoireing: Methods and results. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2020. 2, 7

[23] Shanxin Yuan, Radu Timofte, Gregory Slabaugh, and Ales Leonardis. Aim 2019 challenge on image demoireing: Dataset and study. *IEEE International Conference on Computer Vision (ICCV) Workshops*, 2019. 5

[24] Matthew D Zeiler, Dilip Krishnan, Graham W Taylor, and Rob Fergus. Deconvolutional networks. In *2010 IEEE Computer Society Conference on computer vision and pattern recognition*, pages 2528–2535. IEEE, 2010. 4

[25] 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*, 2017. 5

[26] Yulun Zhang, Kunpeng Li, Kai Li, Lichen Wang, Bineng Zhong, and Yun Fu. Image super-resolution using very deep residual channel attention networks. 2018. 2, 3, 4, 5