C3Net: Demoiréing Network Attentive in Channel, Color and Concatenation

Sangmin Kim[1] Hyungjoon Nam[1] Jisu Kim[1] Jechang Jeong[2*]
[1]Department of Electronics and Computer Engineering, Hanyang University, Seoul, Korea
[2]Department of Electronic Engineering, Hanyang University, Seoul, Korea
{ksmh1652, 013107nam, jisus.kim189}@gmail.com, jjeong@hanyang.ac.kr

Abstract

Attentive neural networks for image restoration are in the spotlight because they got remarkable results both qualitatively and quantitatively. Networks attentive in RGB channels were effective in fields such as single image super-resolution and RAW to RGB mapping. In addition, networks attentive in positions of pixels were used in image denoising. However, networks attentive in positions of pixels, so called spatial attention or pixel attention algorithm, were not as effective in image restoration because the number of pixels in patches of an image is so many that the weights by sigmoid function are insignificant. Also, networks attentive in positions of pixels were mainly used in high-level vision such as image classification and image captioning where there is no need to restore an image itself.

In this paper, we propose a demoiréing network attentive in channel, color, and concatenation, named C3Net. The proposed algorithm uses residual blocks attentive in RGB channels to take advantage of channel attention algorithm. In addition, we introduce a L1 color loss for demoiréing to solve moiré patterns caused by color-striped patterns. Also, we transferred multi-scale information by concatenation, not multiplying with the insignificant weights by sigmoid function. As a result, our proposed C3Net showed state-of-the-art results in the benchmark dataset on NTIRE 2020 demoiréing challenge.

1. Introduction

Moiré patterns are irregular patterns that is formed when a striped or gridded pattern and its displaced version overlap. The displacement of patterns can be the case that the two patterns are not identical; one of the patterns may be moved horizontally, vertically or rotated. In addition, depending on the media displaying those patterns, the printed dots by a printer or the shape of the light sensors in photographing equipment can raise moiré patterns.

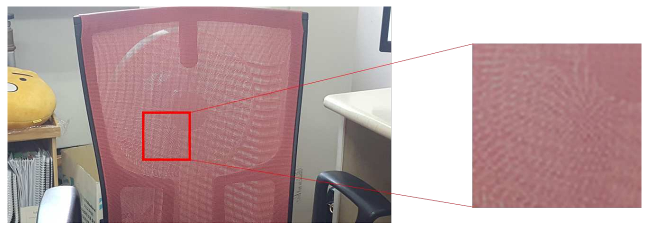

Figure 1. Moiré patterns made by a stripped backrest of a chair and a rotating fan (shot with Galaxy S6).

Consequently, the displacement of patterns creates moiré patterns with frequencies present in neither of the original patterns. These patterns were used for measuring 3-D objects which have shadow moiré patterns [3, 18], sensor for slope movement sensing [15, 30] or even designing [31] because the displacement happens not only in 2-D environment but also in 3-D. Moiré patterns vary from not only where the object with the patterns is but also the position where the object is photographed.

However, in terms of image restoration, moiré patterns can make people who see the patterned images dizzy and degrade the perfomance of images such as peak signal-to-noise ratio (PSNR) and structural similarity (SSIM). To solve these problems, Abraham [1] proposed detecting moiré patterns using multi-level wavelet decomposition and deep learning. Also, algorithms to reduce moiré patterns were proposed according to the media where the images come from. Sasaki et al. [24, 25] pointed out that color moiré patterns on the 3D image can be visual obstacles for viewers and solved it by optical wobbling device or lens shift method in hardware. Hauke et al. [10] tried reconstructing image with moiré patterns using acquired phase-stepping data and discrete Fourier transform [28] and showed good results in X-ray image.

Nowadays, removing moiré patterns are achieved using deep learning because restoring image which has irregular moiré patterns with only hardware or functions is limited. For trained datasets, deep convolutional neural networks (CNN) show remarkable results. Grundhöfer et al. [9]

*Jechang Jeong is the corresponding author.

also tried demoiréing in photographs using CNN with L1 loss, SSIM loss and L2 regularization. Sun et al. [27] used CNN under conditions of diverse resolutions and added all and successfully removed moiré patterns in photographs. Benchmarking CNN under conditions of diverse resolutions, Gao et al. [8] proposed Multi-Scale Feature Enhance (MSFE) and Feature Enhancing Branch (FEB) where Low-level feature and upscaled High-level feature are multiplied. He et al. [11] proposed MopNet which uses U-net [23] structure in Multi-scale Aggregation and Channel-wise Target Edge Predictor. MopNet also utilized Attribute-aware Classifier to categorize the types of moiré patterns and the information of the types of patterns consequently helped to demoiré. Yue et al. [42] proposed a convolutional neural Network with Additive and Multiplicative modules (AMNet) which has a U-net structure and uses residual learning [12] and atrous spatial pyramid pooling (ASPP) [6]. AMNet also benchmarked channel attention [45] in its Channel-wise Multiplicative block to pay attention to color moiré patterns. Yue et al. also compared the results depending on the combination of loss such as MSE loss, adversarial loss [19], and perceptual loss [14].

Shortly, to remove moiré patterns and restore the original images, conventional algorithms focused on three factors: channel-wise attention, customized loss, and multi-scaled information. However, MopNet and AMNet used channel-wise attention only partly. Also, most conventional algorithms used L1 loss suitable for image restoration [49]. In addition, most conventional algorithms are multi-path networks or benchmark U-net for getting the types of moiré patterns in multi-scaled information.

In this paper, we propose a demoiréing network attentive in channel, color, and concatenation, named C3Net. The proposed C3Net tried to complement three factors to facilitate better results. First, C3Net uses residual blocks which consist of convolutional layers and channel attention all over the network. Second, we introduce L1 color loss which is an addition of conventional L1 loss and L1 loss of U and V channels. Lastly, we put a U-net on a block for multi-scaled information and concatenated the multi-scaled information to feature maps followed by residual blocks.

2. Related works

2.1. How to pay attention to moiré patterns

Attentive neural networks, whose outputs depend on the weights of RGB channels, were first introduced in natural language processing (NLP) such as machine translation [4] and text classification [35, 50]. The purpose of the attention algorithm in those networks was to calculate weights among words in a sentence and reflect the weights on the outputs of networks. As a result, many networks showed good results in NLP.

Nowadays, attentive neural networks utilizing images are in the spotlight. As attentive neural networks for NLP calculated weights among words in a sentence, attentive neural networks utilizing images calculated weights among channels in an image. In single image super-resolution (SISR), Zhang et al. [45] introduced this algorithm named channel attention and got better results than any other state-of-the-art algorithms. Also, to solve color-related problem, Uhm et al. [32] proposed W-Net which is a set of two U-net differently trained. W-net used channel-attentional convolution (CA-Convs) block that consists of three sets of a convolutional layer and a LeakyReLU layer and channel attention module and mapped raw image to RGB image successfully.

Attention to positions of pixels of an image, so called spatial attention or pixel attention algorithm, is also a method to give more weights to pixels of the image for desired results. After extracting feature maps, the feature maps pass sigmoid function and are mapped in a range between 0 and 1. Multiplying those weights informs the whole network of how to learn to get better results. Zhang et al. [46] used sigmoid function to feature maps which passed encoder-decoder-shaped mask branch and multiplied the spatial attention maps to feature maps which passed trunk branch with simplified residual blocks [20].

In image captioning, Chen et al. [5] made multi-layers feature maps with good quality by multiplying channel-wise attention weights and spatial attention weights. In image classification, Park et al. [22] introduced bottleneck attention module (BAM) which makes its own BAM attention maps by adding channel attention maps and spatial attention maps. Woo et al. [36] also used a module named convolutional block attention module (CBAM) which makes refined feature maps utilizing channel attention module and spatial attention module. The point that the input feature maps of modules were refined by average pooling and maxpooling also helped the network for learning.

In this paper, C3Net only uses channel attention module in each residual block all over the proposed algorithm, which is different from MopNet [11] and AMNet [42]. The most important reason is that removing moiré patterns is a color-related problem, so the information helpful for demoiréing is likely in RGB channels.

Also, spatial attention module can confuse demoiréing network because the datasets by NTIRE 2020 Demoiréing Challenge [40] are from repeated color-striped patterns such as electronic displays and clothes different from those by AIM 2019 Challenge on Image Demoiréing [41]. Therefore, the weights from spatial attention algorithm become insignificant.

2.2. Color-related loss

When approaching color-related problem, loss for back-propagation was a core factor in image restoration. In case of image colorization, Zhang et al. [44] made the network learn a probability distribution over possible colors using multinomial cross entropy loss. Zhang et al. classified pixels into 313 colors, calculated multinomial cross entropy loss, and combined the loss with MSE Loss. In addition, applying image downscaling in channel-wise, Kim et al. [16] utilized an autoencoder and used a fused L1 loss comparing output with ground truth (GT) both in YUV space of colored pair and Y space of grayscaled pair. One loss is from GT in YUV space and task-aware upscaled image to YUV space, and the other loss is from grayscaled image and task-aware downscaled image to Y space. Kim et al. also adjusted the weights between two losses and showed great colorized results by adding two losses.

In case of RAW to RGB mapping, W-Net used their own color loss which measures the cosine distance between RGB space of predicted image and GT. Uhm et al. defined three values of RGB space of predicted image and GT to each 3-D vector and calculated the cosine distance by divide the inner product of the vectors into the product of the sizes of the vectors. Many conventional algorithms tried to analyze the color space depending on the tasks related to color such as image colorization, RAW to RGB mapping, and demoiréing where Yang et al. [37] made a successful approach by splitting the channels of image into RGB or YUV each to apply layer decomposition on polyphase components (LDPC).

In our proposed algorithm, we include color loss term using L1 loss terms of U and V channels. While conventional L1 loss term of RGB channels helps the network keep the content of the original image, color loss helps the network to restore color-related values whose information is concentrated in U and V channels.

2.3. Concatenation of multi-scaled information

Facilitating better results in image restoration using multi-scaled information was first introduced in U-net [23]. The core idea of U-net is to downscale feature maps, process convolutional layers, concatenate feature maps before downscaling to upscaled feature maps. These contributions make the result show better by multi-scaled information. Conventional algorithms take the advantage of overall U-net in image restoration such as image denoising [21], RAW to RGB Mapping [32], and demoiréing [42].

Multi-scaled information is also well used in multi-path networks. In case of U-net, downscaled feature maps are concatenated on their way upscaling subsequently. However, in case of multi-path networks, the process progresses with the feature maps downscaled. After that, diversely downscaled and processed feature maps gather by residual learning [12] or concatenation. Cheng et al. [7] proposed Multi-scale convolutional network with Dynamic feature encoding for image DeMoiréing (MDDM). MDDM also used adaptive instance normalization (AdaIN), channel attention module, and non-local block [34] in each branch to get good multi-scaled and attentive in channel-wise feature maps. Downscaling the feature maps from x2 to x32, MDDM not only got multi-scaled information, but also detected diverse types of moiré patterns to restore.

Several networks use U-net as a part of networks for multi-scaled information. Wang et al. [33] introduced soft mask branch and the branch downscales and upscales the feature maps by maxpooling and interpolation for image classification. In image denoising, Zhang et al. [46] also use mask branch which downscales and upscales the feature maps by convolutional layers and calculate the weights among pixels of an image, which is named spatial attention. In a different way, Yu et al. [39] used U-net as a block and adopted repeated structure of proposed block. At the same time, the content of the image was kept by residual learning.

In C3Net, we set each block with trunk branch and mask branch. Trunk branch conveys the information of the input image using residual blocks with channel attention modules. Mask branch yields multi-scaled information using a U-net in a block. While Zhang et al. [46] used spatial attention algorithm and multiplied outputs from trunk branch and those from mask branch, C3Net concatenate outputs from trunk branch and those from mask branch to use diverse information to keep not only the content of original image, but also multi-scaled information. In addition, C3Net concatenates outputs from each residual blocks or attention blocks in parallel and outputs from trunk branch and those from mask branch again while Yu et al. [39] conveyed the content of original image using only residual learning and extract multi-scaled feature maps from U-net-based down-up blocks (DUBs) in series. Concatenating many times and in parallel makes C3Net get at moiré patterns and restore the degraded images.

3. Proposed algorithm

3.1. The architecture of the proposed algorithm

Figure 2 shows the entire proposed algorithm, named C3Net for demoiréing in Track 1: Single Image. Keeping the number of channels of feature maps 64, the feature maps pass Attention Via Concatenation blocks (AVCblocks). Also, the feature maps are concatenated every time the maps pass one block, which is first introduced global feature fusion (GFF) by Zhang et al. [47, 48]. The architecture of an AVCblock in Figure 3 is similar to residual non-local attention block (RNAB) [46]; however, the difference from RNAB is that both branches, trunk branch and mask branch, are reinforced by GFF. Also, in case of trunk branch, we

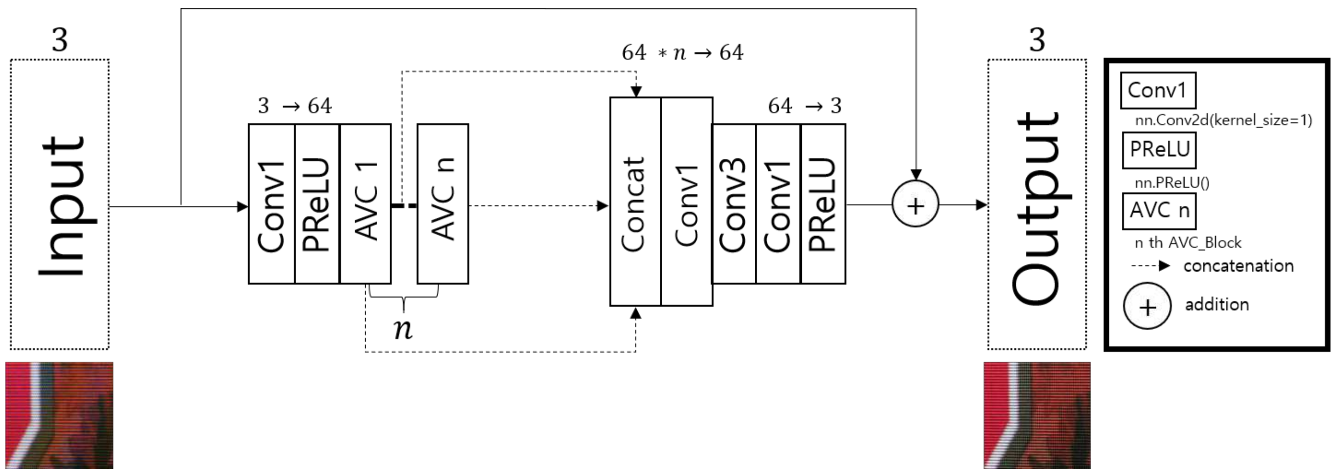

Figure 2. The architecture of C3Net.

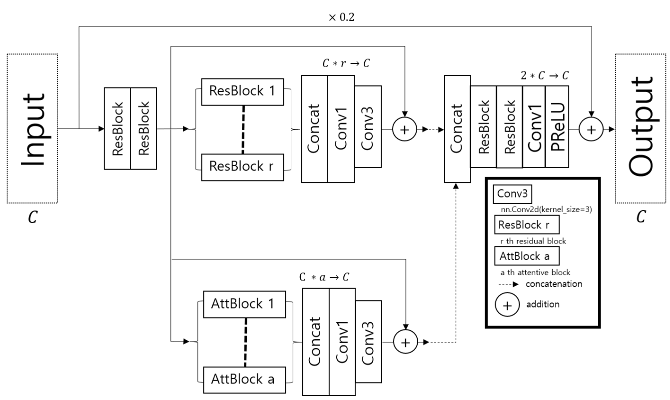

Figure 3. The architecture of Attention Via Concatenation block (AVCblock).

connected residual blocks (ResBlocks) with channel attention module in parallel to extract diverse features into feature maps. In case of mask branch, we connected attentive blocks (AttBlocks) in parallel for the same purpose. ResBlock has a fused structure of residual blocks from Yu et al [38] impressed by Lim et al. [20] and channel attention module [45]. AttBlock in Figure 4 is a U-net [23] of Resblocks, downscaling convolutional layer with activation function, parametric ReLU (PReLU) [13] and upscaling subpixel layer [26] with activation function PReLU. In Figure 3, the number of ResBlocks for each AVCblock, r, is 2 and the number of AttBlocks for each AVCblock, a, is 1 for preventing C3Net from retaining too much parameters.

As the number of blocks and parameters of C3Net increase, the loss is easy to overshoot, so we applied residual scaling by Lim et al. [20] and the scaling factor is 0.2. In addition, residual learning [12] is applied to the outputs of trunk branch and mask branch for the same purpose.

3.2. L1 color loss for demoiréing

In image restoration, it is well-known that L1 loss is suitable for the quality of the results [49]. However, color-related term should be included to focus on removing moiré patterns. Therefore, we propose L1 color loss for demoiréing.

$$L_1 = \| \hat{y}_{RGB} - y_{RGB} \|_1 \tag{1}$$

$$L_{color} = \| \hat{y}_U - y_U \|_1 + \| \hat{y}_U - y_U \|_1 \tag{2}$$

In case of conventional L1 loss, the loss is calculated using RGB channels of an image. We introduce color loss term calculated using U and V channels of the image. While RGB channels include color-related information each, YUV channels include color-related information only in U and V channels; therefore, it is easier to focus on color-related problem. We set the weights of two losses same and the final loss can be written as (3):

$$L_{final} = \frac{L_1 + L_{color}}{2} \tag{3}$$

3.3. Particularities for Track 2: Burst

The most difficult challenge in Track 2 is that input images have zero pixels and random moiré patterns. Originally, the four vertices of images are fixed because images in the datasets from NTIRE 2020 Demoiréing Challenge [40] are square whose length of one side is 128. However, in Track 2, the four vertices of images are not fixed, so

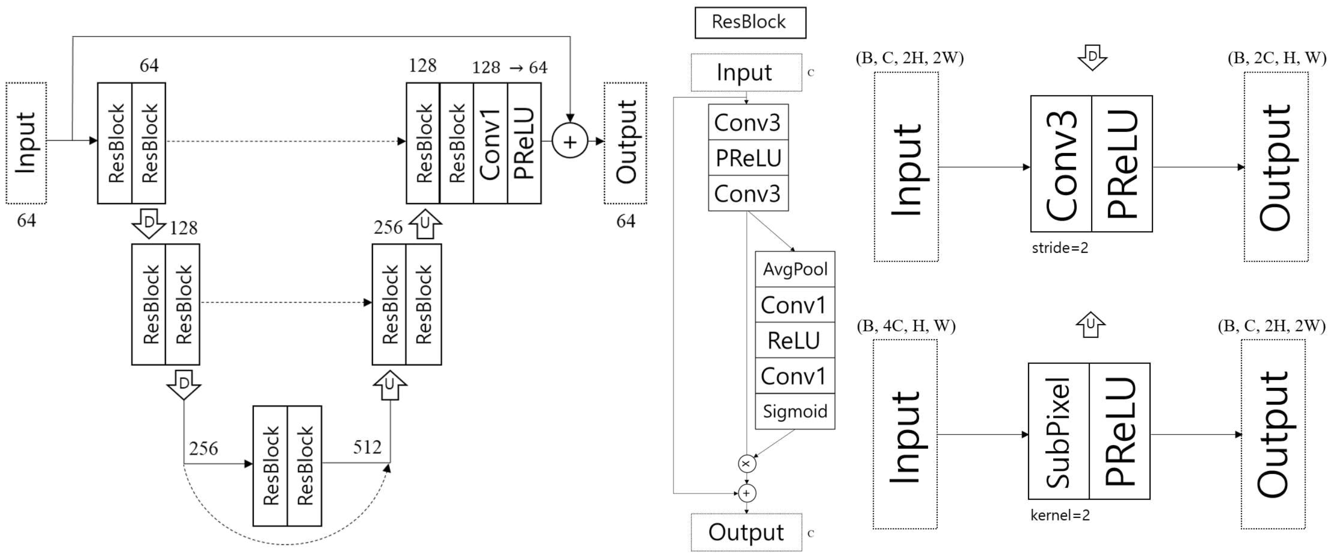

Figure 4. The architecture of an attentive block (AttBlock).

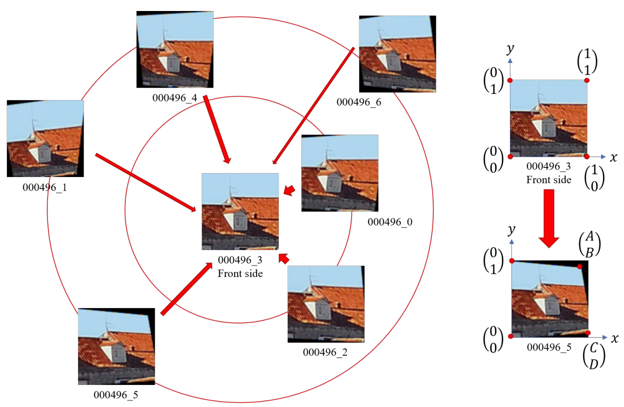

Figure 5. The analysis on the datasets for Track 2 from point of view (left) and linear transformation by matrix (right).

the images seem to be rotated and seen not from the front side where we see. It can be guessed that the images are linearly transformed to make images from different and random point of view. For Track 2, it is important for C3Net to utilize information from different point of view fully. The example and analysis of the datasets are in Figure 5. The circles represent approximate distance from the image photographed from front side. The arrows whose length also represents approximate distance point out the image photographed from front side from the position where the distorted image photographed.

Training CNN, randomly zero pixels hinder learning because the weights in filters of convolutional layers splash where the value of pixels is zero. To solve the problem, we applied chroma key which means an algorithm to fuse two images or videos in the same screen. The method is primarily used when one of the images or videos has monochromatic background. In this paper, we preprocess input images using chroma key with threshold value set to 50. We set the threshold value because zero value makes the image have exposed boundaries. The entire C3Net for Track 2: Burst (C3Net-Burst) is on Figure 6.

The AVCblock for Track 2: Burst (AVC-B block) in Figure 7 is different from that for Track 1: Single Image because the number of input images is 7 times more than that of Track 1. Therefore, the number of AVCblocks is reduced to process C3Net and there is no need to use additional residual learning and residual scaling. In addition, we used global maxpooling [2] in every AVC-B block for getting the best features among 7 images.

4. Experimental results

4.1. Training details

According to the requirements of New Trends in Image Restoration and Enhancement workshop and challenges on image and video restoration and enhancement (NTIRE 2020) Demoiréing Challenge [40], we used 10000 training images, 500 validation images, and 500 testing images whose patch size is all 128 x 128. In Track 1: Single Image, we trained our proposed C3Net using 9000 training images. In Track 2: Burst, we trained our proposed C3Net-Burst using 9999 training sets of 7 burst images. When testing the proposed network, we should set the height and width of an image to multiples of four because of two times of downscaling in U-net [23] structure. In case of datasets for the challenge, the process can be skipped.

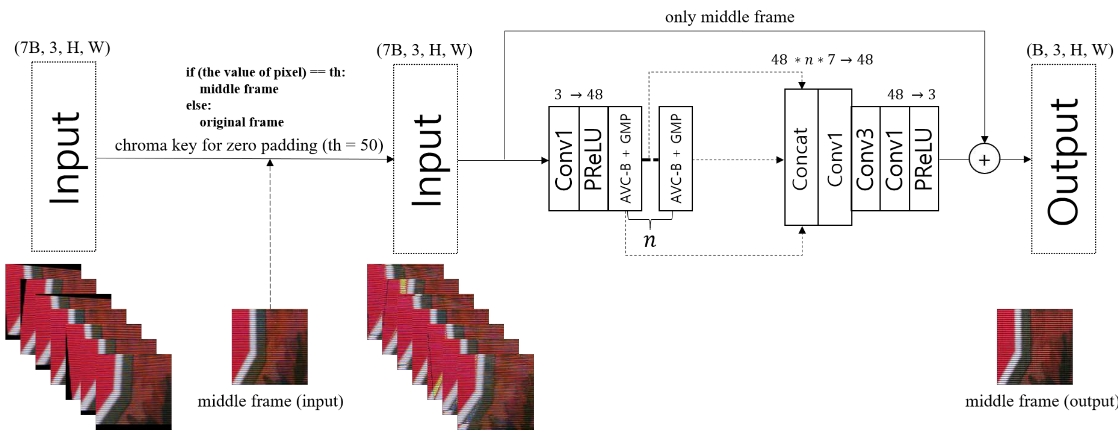

Figure 6. The architecture of C3Net-Burst.

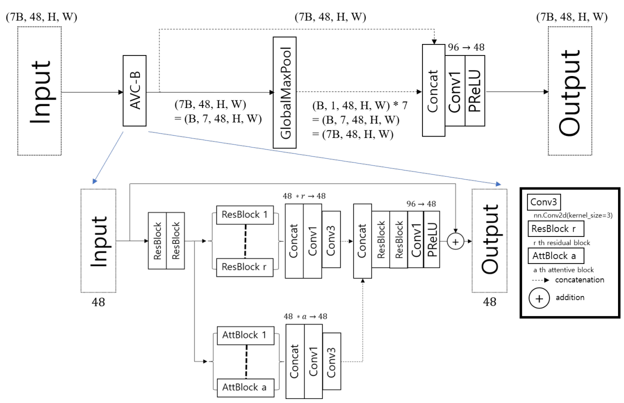

Figure 7. The architecture of AVC-Burst block with global max-pooling (AVC-B + GMP block).

We also trained flipped training images by numpy.flipud() and numpy.fliplr(), which means that the total number of training images quadruples. By training flipped training images, the results can be improved using self-ensemble. In comparison with the state-of-the-art in Table 1, C3Net+ means C3Net with self-ensemble. We trained the network using the datasets whose batch size is. The optimizer we used is ADAM optimizer [17] whose momentum is 0.9 and learning rate is 1e-4. The learning rate halves every 10 epochs, same as about 10k iterations. It took 2.5 hours per epoch and we trained the proposed algorithm about 10 days. Finally, we set the loss to L1 loss or L1 color loss and compare the results from L1 loss with those from L1 color loss.

4.2. Comparison with the state-of-the-art

We compare results from our proposed C3Net with moiré-patterned images, demoiréd images from Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising (DnCNN) [43], Image Super-Resolution via Deep Recursive Residual Network (DRRN) [29], Image Super-Resolution Using Very Deep Residual Channel Attention Networks (RCAN) [45], and the prototypes of C3Net whose performances are verified in the validation phase of the challenge. The conventional algorithms were selected because the algorithms were designed for image restoration and the performances are highly regarded. For fair comparison, All the conventional algorithms were trained in 9999 training images and validated in 500 validation images.

The performance of them is measured in PSNR and SSIM by the validation server of the challenge. Bolded results are the best results. The PSNR results between conventional algorithms, C3Net, and C3Net+ are shown in Table 1. Block means the type of blocks the algorithm used. Conventional algorithms used the block same as the references except upscaling layers in algorithms for SISR such as DRRN and RCAN.

DRRN shows better results than DnCNN because of many parameters and deployment of convolutional layers in close order. RCAN shows near results to C3Net and it is assumed that the reason is that channel attention module in RCAN worked. L1 color loss also helped the conventional algorithms learn demoiréing. Except C3Net+, the proposed algorithms used AVC-B block because AVCblock is designed from AVC-B block to avoid overshooting and applied additional residual learning [12] and residual scaling [20]. N means the number of blocks and more blocks make better results. Trying many experiments, we used L1 loss or L1 color loss. Provided that n is same, using L1 color loss makes better results in validation datasets. Based on these results in Track 1, we set the loss of C3Net-Burst to L1 color loss. Finally, we applied self-ensemble to C3Net+ and got 0.31 more PSNR in validation datasets. The qualitative comparison can be seen at Figure 8.

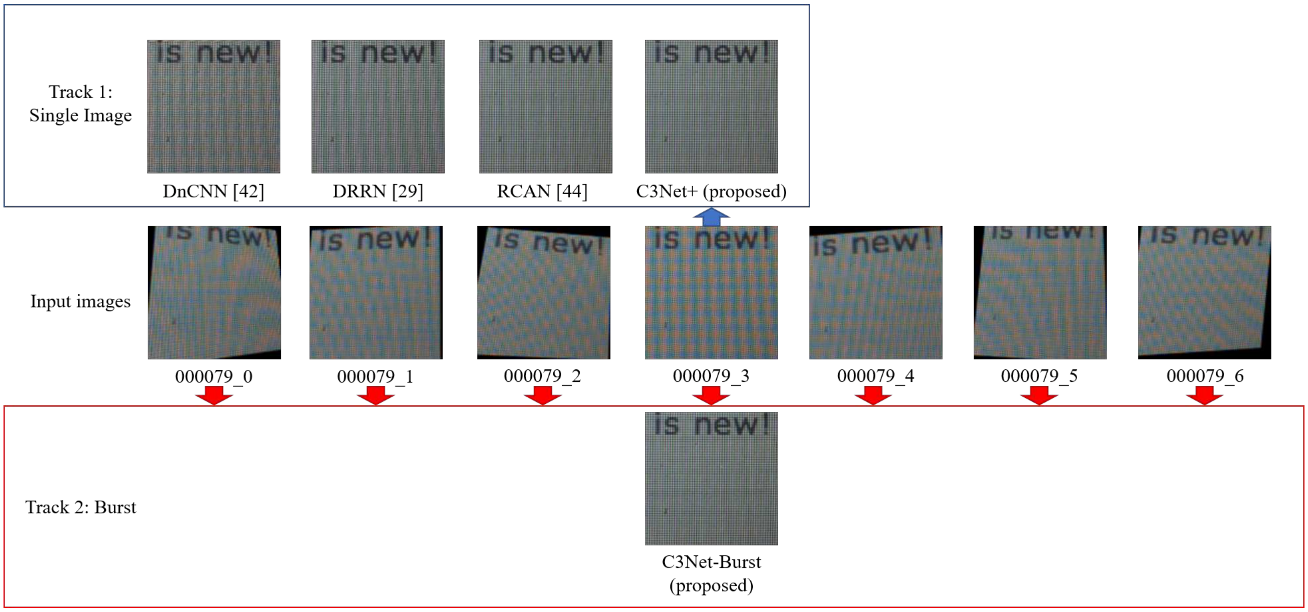

Figure 8. Demoiréing results of conventional algorithms, C3Net+ for Track 1, and C3Net-Burst for Track 2 in validation phase of NTIRE 2020 Demoiréing Challenge.

Track 1	Block	n	# of feature maps	Loss	Ensemble	PSNR
DnCNN	[42]	[42]	64	L1 color	-	28.06
DRRN	[29]	[29]	128	L1 color	-	34.45
RCAN	[44]	[44]	64	L1 color	-	40.25
C3Net	AVC-B	7	64	L1	-	40.80
C3Net	AVC-B	13	64	L1 color	-	41.05
C3Net	AVC	22	64	L1	-	40.99
C3Net+	**AVC**	**22**	**64**	**L1**	**flip** ($\times$**4**)	**41.30**
Track 2	Block	n	# of feature maps	Loss	Ensemble	PSNR
C3Net-Burst	**AVC-B + GMP**	**5**	**48**	**L1 color**	**-**	**40.55**

Table 1. PSNR (dB) comparison of conventional and proposed algorithms on demoiréing.

4.3. NTIRE 2020 Demoiréing Challenge

The proposed algorithm was researched for NTIRE 2020 Demoiréing Challenge [40]. The challenge consists of two tracks: Track 1 for single image and Track 2 for burst images. In case of Track 1: we used AVCblock with additional residual learning and residual scaling, the number of blocks is 22, and used L1 loss for stability. In case of Track 2: we used AVC-B block with global maxpooling, the number of blocks is 5, and used chroma key for removing zero pixels and focusing on removing moiré patterns. In AVC-B block, there is no residual learning and residual scaling because of small number of AVC-B blocks.

In Table 2, our proposed algorithm ranked fourth among 18 entries in Track 1 and among 8 entries in Track 2. Considering that few teams exhibited good results on both tracks, C3Net can be assessed to demoiré admirably.

The key point of NTIRE 2020 Demoiréing Challenge between two tracks is to get better results by fully utilizing 7 burst images. In case of results in Table 2, the results are best results to get from one GPU, NVIDIA Geforce RTX 2080Ti 11GB. Because the size of input images in Track 2 is 7 times more than Track 1, the model size is limited for C3Net-Burst. Therefore, we posted the results under the same conditions of the number of proposed blocks, the number of feature maps, and loss for backpropagation in Table 3. C3Net for Track 1 showed better results than RCAN and C3Net-Burst for Track 2 showed better results than any other three algorithms. Finally, the proposed algorithm exhibited desired results that NTIRE 2020 Demoiréing Challenge required, which is the algorithm to remove random moiré patterns using single image or 7 burst images fully.

Track 1: Single Image				
Rank	Team	Model	PSNR	SSIM
1	1^{st}	1^{st}	42.14	0.99
2	2^{nd}	2^{nd}	41.95	0.99
3	3^{rd}	3^{rd}	41.84	0.99
4	**Reboot**	**C3Net+**	**41.11**	**0.99**
5	5^{th}	5^{th}	41.04	0.99
Track 2: Burst				
Rank	Team	Model	PSNR	SSIM
1	1^{st}	1^{st}	41.95	0.99
2	2^{nd}	2^{nd}	41.88	0.99
3	3^{rd}	3^{rd}	40.64	0.99
4	**Reboot**	**C3Net-Burst**	**40.33**	**0.99**
5	5^{th}	5^{th}	39.05	0.99

Table 2. PSNR (dB) and SSIM comparison of algorithms in testing phase of NTIRE 2020 Demoiréing Challenge.

Algorithm	Block	PSNR
RCAN	[44]	40.25
C3Net	AVC-B	40.31
C3Net-Burst	**AVC-B + GMP**	**40.55**

Table 3. PSNR (dB) comparison between tracks in validation phase of NTIRE 2020 Demoiréing Challenge (n = 5, # of feature maps = 48, Loss = L1 color).

5. Conclusion

Participating in the challenge, we propose an attentive network for removing moiré patterns in channel, color, and concatenation, named C3Net. By adding channel attention module in all over the residual blocks, the proposed network is attentive to channels and easy to solve color-related problems such as moiré patterns. In addition, by introducing L1 color loss, we made a suitable loss fusing L1 loss for image restoration and color loss using U and V channels for color-related problems. We also insist that changing the weights of two losses depending on the tasks in image restoration would bring different results. Lastly, by adopting GFF, U-net in a block and concatenating multi-scaled feature maps, the proposed C3Net can learn diverse types of moiré patterns and handle them. Self-ensemble by training flipped images is also applied for the same purpose. C3Net ranked fourth in both tracks of NTIRE 2020 Demoiréing Challenge [40]. We researched C3Net under the limitations of both tracks and it was possible for C3Net to solve color-related problem and randomly zero pixels in burst images as well. With some variations, the C3Net is expected to showed good results in image restoration and even in other fields related to image. The code and trained models are included in `https://github.com/bmycheez/C3Net`.

Acknowledgement

This work was supported by the fund of On-dream Impact supervised by Hyundai Motor Chung Mong-Koo Foundation of Korea.

References

[1] Eldho Abraham. Moiré pattern detection using wavelet decomposition and convolutional neural network. In *SSCI 2018*. 1

[2] Miika Aittala and Frédo Durand. Burst image deblurring using permutation invariant convolutional neural networks. In *ECCV 2018*. 5

[3] Ali Jamali Avilaq and Amir Hossein Rezaie. 3d face reconstruction using modified shadow moiré. In *ICEE 2013*. 1

[4] Dzmitry Bahdanau, KyungHyun Cho, and Yoshua Bengio. Neural machine translation by jointly learning to align and translate. In *ICLR 2015*. 2

[5] Long Chen, Hanwang Zhang, Jun Xiao, Liqiang Nie, Jian Shao, Wei Liu, and Tat-Seng Chua. Sca-cnn: Spatial and channel-wise attention in convolutional networks for image captioning. In *CVPR 2017*. 2

[6] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L. Yullie. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(4):834–848, 2017. 2

[7] Xi Cheng, Zhenyong Fu, and Jian Yang. Multi-scale dynamic feature encoding network for image demoireing. In *ICCVW 2019*. 3

[8] Tianyu Gao, Yanqing Guo, Xin Zheng, Qianyu Wang, and Xiangyang Luo. Moiré pattern removal with multi-scale feature enhancing network. In *ICMEW 2019*. 2

[9] Anselm Grundhöfer and Gerhard Röthlin. Camera-specific image quality enhancement using a convolutional neural network. In *ICIP 2017*. 1

[10] C. Hauke, M. Leghissa, T. Mertelmeier, M. Radicke, S. Sutter, T. Weber, G. Anton, and L. Ritchl. Moiré artefact reduction in talbot-lau x-ray imaging. In *ISBI 2018*. 1

[11] Bin He, Ce Wang, Boxin Shi, and Ling-Yu Duan. Mop moiré patterns using mopnet. In *ICCV 2019*. 2

[12] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *CVPR 2016*. 2, 3, 4, 6

[13] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *ICCV 2015*. 4

[14] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In *ECCV 2016*. 2

[15] Isabelle Chan Shieu Juinn, Mani Maran Ratnam, and Fauziah Ahmad. Novel moiré-pattern based tilt sensor for slope movement sensing: Repeatability study. In *ICCSCE 2014*. 1

[16] Heewon Kim, Myungsub Choi, Bee Lim, and Kyoung Mu Lee. Task-aware image downscaling. In *ECCV 2018*. 3

[17] Diederik P. Kingma and Jimmy Lei Ba. Adam: A method for stochastic optimization. In *ICLR 2015*. 6

[18] Yun-Long Lay, Hui-Jen Yang, Chern-Sheng Lin, and Wei-Yu Chen. 3d object measurement by shadow moiré. In *OPTICS 2010*. 1

[19] Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, and Wenzhe Shi. Photo-realistic single image super-resolution using a generative adversarial network. In *CVPR 2017*. 2

[20] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. Enhanced deep residual networks for single image super-resolution. In *CVPRW 2017*. 2, 4, 6

[21] Bumjun Park, Songhyun Yu, and Jechang Jeong. Densely connected hierarchical network for image denoising. In *CVPRW 2019*. 3

[22] Jongchan Park, Sanghyun Woo, Joon-Young Lee, and In So Kweon. Bam: Bottleneck attention module. *arXiv preprint arXiv:1807.06514*, 2018. 2

[23] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *MICCAI 2015*. 2, 3, 4, 5

[24] Hisayuki Sasaki, Naoto Okaichi, Hayato Watanabe, Masanori Kano, Masahiro Kawakita, and Tomoyuki Mishina. Color moiré reduction and resolution enhancement technique for integral three-dimensional display. In *3DTV-CON 2017*. 1

[25] Hisayuki Sasaki, Hayato Watanabe, Naoto Okaichi, Kensuke Hisatomi, and Masahiro Kawakita. Color moiré reduction method for thin integral 3d displays. In *IEEE VR 2019*. 1

[26] 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*. 4

[27] Yujing Sun, Yizhou Yu, and Wenping Wang. Moiré photo restoration using multiresolution convolutional neural networks. *IEEE Transactions on Image Processing*, 27(8):4160–4172, 2018. 2

[28] D. Sundararajan. *The discrete Fourier transform: theory, algorithms and applications*. World Scientific, 2001. 1

[29] Ying Tai, Jian Yang, and Xiaoming Liu. Image super-resolution via deep recursive residual network. In *CVPR 2017*. 6

[30] P. Y. Tan, K. S. Yen, M. M. Ratnam, and F. Ahmad. Novel moiré-pattern based tilt sensor for slope movement sensing: Repeatability study. In *TENCON 2016*. 1

[31] Pei-Hen Tsai and Yung-Yu Chuang. Target-driven moiré pattern synthesis by phase modulation. In *ICCV 2013*. 1

[32] Kwang-Hyun Uhm, Seung-Wook Kim, Seo-Won Ji, Sung-Jin Cho, Jun-Pyo Hong, and Sung-Jea Ko. W-net: Two-stage u-net with misaligned data for raw-to-rgb mapping. In *ICCVW 2019*. 2, 3

[33] Fei Wang, Mengqing Jiang, Chen Qian, Shuo Yan, Cheng Li, Honggang Zhang, Xiaogang Wang, and Xiaoou Tang. Residual attention network for image classification. In *CVPR 2017*. 3

[34] Xiaolong Wang, Ross Girshick, Abhinav Gupta, and Kaiming He. Non-local neural networks. In *CVPR 2018*. 3

[35] Yequan Wang, Minlie Huang, Li Zhao, and Xiaoyan Zhu. Neural machine translation by jointly learning to align and translate. In *EMNLP 2016*. 2

[36] Sanghyun Woo, Jongchan Park, Joon-Young Lee, and In So Kweon. Cbam: Convolutional block attention module. In *ECCV 2018*. 2

[37] Jingyu Yang, Xue Zhang, Changrui Cai, and Kun Li. Demoiréing for screen-shot images with multi-channel layer decomposition. In *VCIP 2017*. 3

[38] Songhyun Yu and Jechang Jeong. Local excitation network for restoring a jpeg-compressed image. *IEEE Access*, 7:138032–138042, 2019. 4

[39] Songhyun Yu, Bumjun Park, and Jechang Jeong. Deep iterative down-up cnn for image denoising. In *CVPRW 2019*. 3

[40] Shanxin Yuan, Radu Timofte, Ales Leonardis, Gregory Slabaugh, et al. Ntire 2020 challenge on image demoireing: Methods and results. In *CVPRW 2020*. 2, 4, 5, 7, 8

[41] Shanxin Yuan, Radu Timofte, Gregory Slabaugh, and Ales Leonardis. Aim 2019 challenge on image demoireing: Dataset and study. *arXiv preprint arXiv:1911.02498*, 2019. 2

[42] Huanjing Yue, Yan Mao, Lipu Liang, Hongteng Xu, and Chunping Hou. Recaptured screen image demoireing. *IEEE Transactions on Circuits and Systems for Video Technology*, pages 1–1, 2020. 2, 3

[43] Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, and Lei Zhang. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. In *CVPR 2017*. 6

[44] Richard Zhang, Phillip Isola, and Alexei A. Efros. Colorful image colorization. In *ECCV 2016*. 3

[45] Yulun Zhang, Kunpeng Li, Kai Li, Lichen Wang, Bineng Zhong, and Yun Fu. Image super-resolution using very deep residual channel attention networks. In *CVPR 2018*. 2, 4, 6

[46] Yulun Zhang, Kunpeng Li, Kai Li, Bineng Zhong, and Yun Fu. Residual non-local attention networks for image restoration. In *ICLR 2019*. 2, 3

[47] Yulun Zhang, Yapeng Tian, Yu Kong, Bineng Zhong, and Yun Fu. Residual dense network for image super-resolution. In *CVPR 2018*. 3

[48] Yulun Zhang, Yapeng Tian, Yu Kong, Bineng Zhong, and Yun Fu. Residual dense network for image restoration. *arXiv preprint arXiv:1812.10477*, 2018. 3

[49] Hang Zhao, Orazio Gallo, Iuri Frosio, and Jan Kautz. Loss functions for image restoration with neural networks. *IEEE Transactions on computational imaging*, 3(1):47–57, 2016. 2, 4

[50] Peng Zhou, Wei Shi, Jun Tian, Zhenyu Qi, Bingchen Li, Hongwei Hao, and Bo Xu. Attention-based bidirectional long short-term memory networks for relation classification. In *ACL 2016*. 2