NTIRE 2020 Demoireing Challenge Factsheet CubeDemoireNet: Enhanced Multi-Scale Network for Image Demoireing

Xiaohong Liu, Linhui Dai, and Jun Chen

March 25, 2020

1 Team details

- Team name: Mac AI
- Team leader: Xiaohong Liu
- Team leader address, phone number, and email:
  1280 Main Street West, ITB A203, Hamilton, Ontario, Canada; 905-525-9140 ext: 27904; liux173@mcmaster.ca
- Rest of the team members: Linhui Dai and Jun Chen
- Team website URL: N/A
- Affiliation: McMaster University
- Affiliation of the team and/or team members with NTIRE2020 sponsors: N/A
- User names and entries on the NTIRE2020 Codalab competitions: proteus1991
- Best scoring entries of the team during development/validation phase: Track 1: PSNR: 40.82, SSIM: 0.9910; Track 2: PSNR: 40.86, SSIM: 0.9911.
- Link to the codes/executables of the solution(s): Attached in email.
- Link to the restoration results of all frames: Attached in email.

2 Contribution details

- Title of the contribution
  Enhanced Multi-Scale Network for Image Demoireing

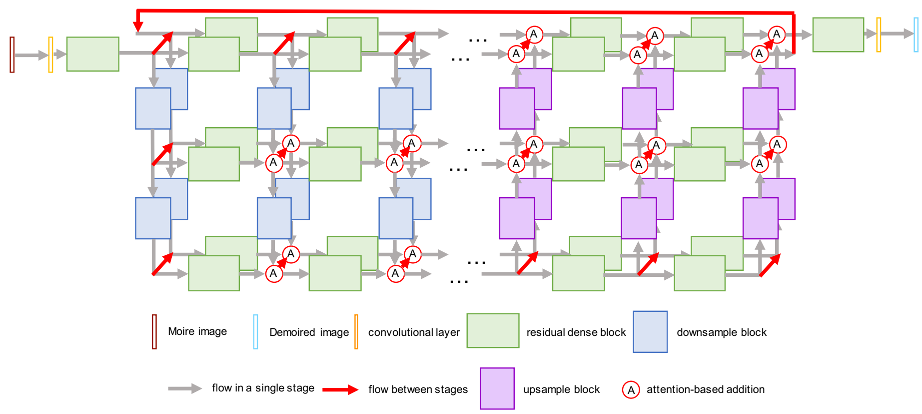

Figure 1: The main architecture of the proposed CubeDemoireNet.

- General method description
  Recently, using cameras or mobile phones to take photographs on digital screens becomes a convenient manner to temporally store the messages that might be accessed later. However, Moiré patterns usually appear in this circumstance, which are caused by the imperfect alignment between the LED lattices of digital screens and the Color Filter Array (CFA) of camera sensors [1]. Due to the fact that this misalignment is quite random and not constrained to a certain pattern, Moiré is capable of embedding in a variety of frequencies with dynamic patterns that is usually difficult to remove. Inspired by [3], we propose a new network, named CubeDemoireNet, which enhances the conventional multi-scale network by 1) facilitating information exchange across different scales and stages to alleviate the underlying bottleneck issue, and 2) employing the attention mechanism to identify the dynamic Moiré patterns, and thus eliminate the ones effectively with the preservation of image texture.

- Description of the particularities of the solutions deployed for each of the challenge competitions or tracks.

  To tackle Moiré effect, recent advances [8] extensively resort to multi-scale structure that can deal with features at various spatial resolutions to accommodate different frequency bands accordingly. Due to the fact that dynamic Moiré patterns usually embeds in a broad range of frequencies, the benefit of multi-scale structure is evident. However, the conventional multi-scale network is designed hierarchically that often suffers the bottleneck issue caused by insufficient information flow among different scales. In addition, except for demoireing [1], other restoration problems including, among others, super-resolution [7] and dehazing [4] advert the manner that gradually recovers the degraded image from coarse one to fine one. Extensive experiments have been conducted and provide the solid justification that this manner can further improve the quality of restored images, even in the one that simply stacks the network many times [1, 7, 4]. Despite the coarse-to-fine manner has been widely employed, most of existing methods stack different stages sequentially. To the best of our knowledge, none of them densely connects the intermediate features between adjacent stages to facilitate the information flow and the speed of convergence. To solve these problems, inspired by [3], we propose a new network, named CubeDemoireNet, which enables to promote information exchange across different scales and stages via dense connections using upsamling/downsampling blocks to alleviate the underlying bottleneck issue prevalently existed in conventional multi-scale network. Moreover, we further integrate the attention mechanism into our network that is capable of preserving image texture while eliminating the dynamic Moiré patterns. Fig. 1 demonstrate the main architecture of the proposed CubeDemoireNet. Our network consists of two convolutional layers and three blocks that are residual dense block (RDB), upsampling block, and downsampling block respectively. More specifically, we adopt the same settings in [9] as the ones in our RDB. The upsampling and downsampling blocks are comprised of only two convolutional layers to adjust the spatial resolutions accordingly. As for attention mechanism, we choose the one employed in SENet [2].

In single and burst demoireing tracks, the proposed CubeDemoireNet is used as the backbone. For clearer explanation, each row (comprised of the RDBs) in our network corresponds to a different scale, each column (comprised of the upsampling/downsampling blocks) serves as a bridge to connect each scale, and each stage consists of the coarse-to-fine manner. Note that the intermediate features from former stage are fed into the same position of the current stage to enhance information exchange, and more importantly, the output of the former stage is the input of the current stage. In our method, the row, column, and stage is set to 3, 12, and 3 respectively for both tracks. Moreover, we found that by adding a refinement network can further improve the demoireing performance. Therefore, a modified U-net [5] 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 [7] is employed. In summary, we use CubeDemoireNet-UNet for Track 1 and PCD-CubeDemoireNet-UNet for Track 2 in this challenge.

3 Global Method Description

- Total method complexity:
  For Track 1, the total parameters are 13,623,878.
  For Track 2, the total parameters are 13,817,466.
- Which pre-trained or external methods / models have been used (for any stage, if any)

The PCD from [7] and a modified UNet inspired by [5].

- Which additional data has been used in addition to the provided NTIRE training and validation data (at any stage, if any)
  None.

- Training description
  Track 1: three individual training steps are used for this track. Firstly, we train our backbone, which is named as the CubeDemoireNet. The batch size is set to 8 and a specific decay strategy is used, where the initial learning rate is set to 2e-4 and decays 0.5 times every 50 epochs for total 400 epochs. Secondly, we add a UNet as the refinement for the output of the backbone, and freeze the backbone weight to train the refinement network individually, where the training strategy keeps the same but only trained for 200 epochs. Finally, we unfreeze the backbone weight and train both of them together. Due to the constrain of the GPU memory, the batch size can only set to 4, and we use a smaller learning rate, 1e-5, to finetune the entire system for other 200 epochs. It is worth noting that employing the refinement network can further improve the overall performance by 0.3 dB as compared to the one without it.

  Track 2: two additional training steps are employed based on the results of track 1 since we regard this track as an extension from the previous one. To deal with a sequence of wrapped images, the PCD alignment from [7] is capitalized on as our framework head. Then, the output of the PCD alignment is fed into the subsequent network. Therefore, we first freeze the backbone and refinement weights and train the PCD alignment 50 epochs, in which the initial learning rate is set to 1e-4 and decayed 0.5 times every 10 epoch. After this step, we train the entire system for other 100 epochs by keeping the same training strategy except for setting the initial learning rate to 1e-5.

- Testing description
  Track 1: Inspired by [6], we use the ensemble method to augment the test data, which produces 7 new variants for each input. Therefore, each demoired image in test data is obtained by taking the average of all outputs processed by the network from augmented inputs.

  Track 2: Similar to Track 1, the ensemble method is adopted for Track 2 as well. Therefore, we augment each burst sequence to in total 8 ones (including itself) to recover the demoired image at center.

- Novelty degree of the solution and if it has been previously published
  Inspired by the GridDehazeNet [3], the proposed CubeDemoireNet first considers the intermediate feature exchanges among different stages to facilitate the information flow and the speed of network convergence, which can be regarded as an enhanced version of previous work.

4 Competition particularities

- Any particularities of the deployed solution for the competition (if applicable)
  N/A

5 Ensembles and fusion strategies

- Describe in detail the use of ensembles and/or fusion strategies (if any).
  The proposed method uses ensembles comprised of in total 8 variants (original, rotated 90°, rotated 180°, rotated 270°, flipped, 90° & flipped, 180° & flipped, and 270° & flipped ones) mentioned in [6].

- What was the benefit over the single method?

  The improvement of employing ensembles is non-trivial. Since the ground-truth of validation data is not released. To evaluate the benefit of ensembles, we extract the last 500 images from the provided training data to validate our methods by calculating the PSNR and SSIM values. In our experiments, non-ensembles version in Track 1 achieves 41.42 dB and 0.9919 for PSNR and SSIM. By employing ensembles, the PSNR and SSIM values increase 0.47 dB and 0.0006 respectively. As for Track 2, we observe the benefit of ensembles is consistent with the one in Track 1.

6 Technical details

- Language and implementation details (including platform, memory, parallelization requirements)
  The language and framework we use to implement the CubeDemoireNet is Python 3.7 and PyTorch 1.1 based on a PC with two NVIDIA GeForce GTX 1080 Ti GPUs and 32GB memory.

- Human effort required for implementation, training and validation?
  None.

- Training/testing time? Runtime at test per image.
  For Track 1, the runtime per image is 0.1635s.
  For Track 2, the runtime per image is 0.2753s.

- Comment the robustness and generality of the proposed solution(s)? Is it easy to deploy it for other sets of downscaling operators?
  The proposed CubeDemoireNet is a fully CNN. Thus, it is quit flexible to fit most of resolutions and operations. From the obtained experimental results, we believe the proposed method is robust while dealing with a variety of Moiré patterns.

- Comment the efficiency of the proposed solution(s)?
  The proposed CubeDemoireNet is relatively efficient as compared to others in validation board.

7 Other details

- Planned submission of a solution(s) description paper at NTIRE2020 workshop.
  If we are invited for a submission, we will write a paper to describe the proposed network in detail.
- General comments and impressions of the NTIRE2020 challenge.
  The NTIRE2020 challenge is fantastic and gives a rare opportunity for who works on the low-level vision area to compete and communicate with other talent researchers.

References

[1] Xi Cheng, Zhenyong Fu, and Jian Yang. Multi-scale dynamic feature encoding network for image demoiréing. *arXiv preprint arXiv:1909.11947*, 2019.

[2] Jie Hu, Li Shen, and Gang Sun. Squeeze-and-excitation networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 7132–7141, 2018.

[3] Xiaohong Liu, Yongrui Ma, Zhihao Shi, and Jun Chen. Griddehazenet: Attention-based multi-scale network for image dehazing. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 7314–7323, 2019.

[4] Wenqi Ren, Lin Ma, Jiawei Zhang, Jinshan Pan, Xiaochun Cao, Wei Liu, and Ming-Hsuan Yang. Gated fusion network for single image dehazing. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3253–3261, 2018.

[5] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 234–241. Springer, 2015.

[6] Radu Timofte, Rasmus Rothe, and Luc Van Gool. Seven ways to improve example-based single image super resolution. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1865–1873, 2016.

[7] Xintao Wang, Kelvin CK Chan, Ke Yu, Chao Dong, and Chen Change Loy. Edvr: Video restoration with enhanced deformable convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2019.

[8] Shanxin Yuan, Radu Timofte, Gregory Slabaugh, Ales Leonardis, Bolun Zheng, Xin Ye, Xiang Tian, Yaowu Chen, Xi Cheng, Zhenyong Fu, et al. Aim 2019 challenge on image demoireing: Methods and results. *arXiv preprint arXiv:1911.03461*, 2019.

[9] Yulun Zhang, Yapeng Tian, Yu Kong, Bineng Zhong, and Yun Fu. Residual dense network for image super-resolution. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2472–2481, 2018.