NTIRE 2021 Challenge on Image Deblurring


Abstract

Motion blur is a common photography artifact in dynamic environments that typically comes jointly with the other types of degradation. This paper reviews the NTIRE 2021 Challenge on Image Deblurring. In this challenge report, we describe the challenge specifics and the evaluation results from the 2 competition tracks with the proposed solutions. While both the tracks aim to recover a high-quality clean image from a blurry image, different artifacts are jointly involved. In track 1, the blurry images are in a low resolution while track 2 images are compressed in JPEG format. In each competition, there were 338 and 238 registered participants and in the final testing phase, 18 and 17 teams competed. The winning methods demonstrate the state-of-the-art performance on the image deblurring task with the jointly combined artifacts.

1. Introduction

Motion blur is a prevalent artifact in dynamic scene photography. Hand-held cameras are prone to shake while the objects in the scene can move during the exposure. Moreover, images are typically degraded from joint visual artifacts including motion blur, low resolution, compression artifacts, noise, etc. Image deblurring aims to recover a clean image from such a degraded blurry image.

Most modern image restoration techniques including image deblurring adopt machine-learning approaches that derive knowledge from training data. For deblurring problem, the pairs of blurry and sharp images could be obtained by synthesizing blur from high-speed videos [60, 73, 64, 69, 59]. Especially, REDS dataset [59] is designed to generate high-quality images as well as realistic image degradation. Recently, there were attempts to construct datasets with real blurry images by using a beam splitter [67, 111] and 2 cameras. For such hardware-based approaches, evenly splitting the brightness and precisely aligning the image pair remains an issue.

To develop and benchmark deblurring algorithms, image and video deblurring challenges were hosted in the NTIRE 2019 and 2020 workshops. In the NTIRE 2019 Challenge,
video deblurring [62] and super-resolution methods under low-resolution [63] were developed. In the NTIRE 2020 Challenge, single image deblurring methods [61] are benchmarked.

Succeeding the prior challenges, NTIRE 2021 Challenge on Image Deblurring considers image deblurring problem under additional artifacts. In track 1, the blurry images are in a lower resolution than the target resolution. Thus, high-frequency information is more scarce in the input. In track 2, the blurry image suffers from JPEG compression artifacts. In contrast to most deblurring methods that only consider pure motion blur, the joint image restoration tasks pose more challenging and practical scenario.

This challenge is one of the NTIRE 2021 associated challenges: nonhomogeneous dehazing [3], defocus deblurring using dual-pixel [1], depth guided image relighting [18], image deblurring, multi-modal aerial view imagery classification [47], learning the super-resolution space [55], quality enhancement of heavily compressed videos [92], video super-resolution [72], perceptual image quality assessment [21], burst super-resolution [6], high dynamic range [66].

2. Related Works

We describe the deep learning based image deblurring methods as well as the super-resolution and image deblocking (decompression).

2.1. Image Deblurring

Deep learning was applied to dynamic scene deblurring by constructing datasets with high-speed cameras [60, 73, 64]. Multi-scale networks [60, 76, 20] followed the coarse-to-fine approaches in optimization based frameworks [40, 12, 89, 32, 33]. Motivated that motion blur is spatially varying, spatially non-uniform operations as well as convolution were adopted. Spatially variant RNN was proposed as a deconvolution operator [103] and deformable convolution [114] was used to approximate the shape of blur kernel [99]. In contrast to the single feed-forward computation, MTRNN [65] proposed to remove partial blur multiple times with a small module. In [69], more attention was paid to human bodies as they tend to be the main objects in photography.

On the other hand, there were efforts to optimize such models with focus on perceptual quality. Adversarial loss [60, 34, 34], perceptual loss [34, 35] were used. Also, unsupervised training with cycle-consistency [54] was attempted in domain-specific deblurring.

Specific to face and text images, [91] proposed a joint deblurring and super-resolution model with deep learning. To cope with the ill-posedness of the joint task, adversarial training framework is employed to learn a category-specific prior. Later, dual branch architectures were proposed [106, 107]. In [106], the features from deblurring module and the super-resolution feature extraction module are fused by gate module to obtain high-resolution reconstruction result. In contrast, [107] uses a feature extraction module followed by the deblurring module and the high-resolution prediction module. The auxiliary deblurring branch is used to aid train the feature extraction module.

On the other hand, little attempts were made to handle compression artifacts in deblurring task. In case of video deblurring, MPEG compression was considered in NTIRE 2019 Challenge on Video Deblurring [62].

2.2. Image Super-Resolution

From the the early CNNs for super-resolution [16, 31], many model architectures were proposed. Faster models were developed using sub-pixel convolutions [17, 70]. Later, residual networks [27] were widely adopted in the later methods [38, 45, 28] as well as dense connections [79, 109]. Multi-scale models were also proposed to handle information in different frequency bands [37, 8]. Back projection networks were developed to provide iterative feedback mechanism [26, 43]. In order to focus on relatively more useful features, attention modules were applied to the channels [108, 13] and spatial location [57] on feature maps. Also, high-level information were jointly used to aid super-resolution performance [83].

In contrast to conventional super-resolution methods considering bicubic downsampling, kernel-based methods tried to handle general downsampling methods [22, 104, 112]. To make deployed super-resolution model adapt to the test image, meta-learning was applied [29].

2.3. Image Deblocking

Early JPEG artifacts reduction mainly relied on image filtering [46, 95], transformed domain [44] or via optimization [93, 42]. Sparsity was exploited for regularization [9, 53, 52]. More recent deep learning methods learn to suppress the artifacts by minimizing reconstruction error on training set [15, 75, 98, 19]. To reflect the compression model of JPEG compression, the loss function were calculated on the frequency domain [23, 24, 94]. The traditional sparse coding schele was reflected in neural networks [19].

3. NTIRE 2021 Challenge

We hosted the NTIRE 2021 Challenge on Image Deblurring in order to encourage the community to develop the state-of-the-art algorithms for dynamic image deblurring in the wild condition. The main objective of the challenge is to handle motion blur under additional joint degradation artifacts. Following the NTIRE 2019 and 2020 challenges [62, 61], we use the REDS dataset [59] to measure the performance of the results.
<table>
<thead>
<tr>
<th>Team</th>
<th>PSNR↑</th>
<th>SSIM↑</th>
<th>LPIPS↓</th>
<th>Runtime</th>
</tr>
</thead>
<tbody>
<tr>
<td>VIDAR</td>
<td>29.04</td>
<td>0.8416</td>
<td>0.2397</td>
<td>1.0</td>
</tr>
<tr>
<td>netal</td>
<td>28.91</td>
<td>0.8246</td>
<td>0.2569</td>
<td>12.4</td>
</tr>
<tr>
<td>NJUST-IMAG</td>
<td>28.51</td>
<td>0.8172</td>
<td>0.2547</td>
<td>6.4</td>
</tr>
<tr>
<td>SRC-B</td>
<td>28.44</td>
<td>0.8158</td>
<td>0.2531</td>
<td>0.9</td>
</tr>
<tr>
<td>Baidu</td>
<td>28.44</td>
<td>0.8135</td>
<td>0.2704</td>
<td>40.8</td>
</tr>
<tr>
<td>MMM</td>
<td>28.42</td>
<td>0.8132</td>
<td>0.2685</td>
<td>14.3</td>
</tr>
<tr>
<td>Imagination</td>
<td>28.36</td>
<td>0.8130</td>
<td>0.2666</td>
<td>7.3</td>
</tr>
<tr>
<td>Noah_CVlab</td>
<td>28.33</td>
<td>0.8132</td>
<td>0.2606</td>
<td>24.5</td>
</tr>
<tr>
<td>TeamInception</td>
<td>28.28</td>
<td>0.8110</td>
<td>0.2651</td>
<td>0.9</td>
</tr>
<tr>
<td>ZOCS_Team</td>
<td>28.25</td>
<td>0.8108</td>
<td>0.2636</td>
<td>2.2</td>
</tr>
<tr>
<td>Mier</td>
<td>28.21</td>
<td>0.8109</td>
<td>0.2646</td>
<td>17.3</td>
</tr>
<tr>
<td>INFINITY</td>
<td>28.11</td>
<td>0.8064</td>
<td>0.2734</td>
<td>2.7</td>
</tr>
<tr>
<td>DMLAB</td>
<td>27.78</td>
<td>0.7960</td>
<td>0.2830</td>
<td>6.5</td>
</tr>
<tr>
<td>RTQSA-Lab</td>
<td>27.64</td>
<td>0.7956</td>
<td>0.2730</td>
<td>1.6</td>
</tr>
<tr>
<td>Yonsei-MCMCL</td>
<td>27.61</td>
<td>0.7936</td>
<td>0.2885</td>
<td>0.3</td>
</tr>
<tr>
<td>SCUT-ZS</td>
<td>27.55</td>
<td>0.7935</td>
<td>0.2785</td>
<td>0.3</td>
</tr>
<tr>
<td>withdrawn team</td>
<td>27.44</td>
<td>0.7902</td>
<td>0.2850</td>
<td>1.0</td>
</tr>
<tr>
<td>bicubic upsampling</td>
<td>24.06</td>
<td>0.6817</td>
<td>0.5120</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 1: NTIRE 2021 Image Deblurring Challenge results measured on the REDS [59] test dataset. Teams are ordered by ranks in terms of PSNR(dB). The running time is the average test time (sec) taken to generate a single output image in reproduction process using 1 Quadro RTX 8000 GPU with 48GB VRAM. We note that the reported timing includes I/O and initialization overhead due to the difficulty in measuring pure model inference time by modifying each implementation.

3.1. Tracks and Competitions

In this challenge, we considered commonly witnessed visual artifacts, low-resolution and the JPEG compression as well as the motion blur. Both the degradations make the removal of motion blur to be more difficult. The competition consists of 2 tracks: (1) Low Resolution (2) JPEG Artifacts.

**Image Deblurring Track 1. Low Resolution** aims to develop single-image deblurring methods under $\times 4$ low resolution image than the target resolution. A joint deblurring and super-resolution task is posed.

**Image Deblurring Track 2. JPEG Artifacts** provides the blurry images under JPEG compression. The images are compressed by $\times 4$ ratio to keep a similar degree of information loss as Track 1.

**Competitions** Both the tracks are hosted on the CodaLab competition platform. Each participant is required to register to the CodaLab challenge tracks to access the data and submit their deblurred results. During the development phase, the participants use their training set to develop solutions. The online feedback on part (every 10th) of the validation data was available. Due to the large size of the validation set, the participants were provided with the validation data ground truth for local evaluation. At the testing phase, each team were required to submit part of the testing set results to the CodaLab server. Parallel with the online submission, all the deblurred images and the inference code was submitted via email.

**Evaluation** The primary evaluation metric in this challenge is PSNR. To supplement and provide additional information, SSIM [87] and LPIPS [105] is also measured. The running time was measured by the organizers with the code provided by the participants, checking the reproducibility of each solution.

4. Challenge Results

Each challenge track had 338 and 238 registered participants. In each track, 18 and 17 teams submitted the results in the final testing phase. The deblurred images were submitted along with the inference code and the trained weights for the organizers to check the reproducibility.

Table 1 shows the measured performance of each team’s solution as well as their inference speed. The inference speed was measured by the organizers in a single platform. We used Intel Xeon Gold 6248 CPU and NVIDIA Quadro RTX 8000 GPU, Samsung 860 EVO 4TB SSD.

* Solutions from DuLang and LAB FUD teams were not reproducible from the submitted code.
4.1. Architectures and Main Ideas

There were a few novel ideas and several shared strategies between the submitted solutions. In track 1, inspired from the video deblurring technique in EDVR [84], VIDAR and Imagination teams used pyramid deformable convolutions to align multiple features from a single image. netai and Noah_CVLab teams used multi-task training method to optimize features for joint deblurring and super-resolution. Transformer architecture [82] was used by Noah_CVLab and ZOCS teams. Non-local module [85] was used by NJUST-IMAG team. In track 2, to overcome the limitation of batch normalization [30], half-instance normalization scheme was proposed by The Fat, The Thin and The Strong team. Object edge information was exploited by Yonsei-MCML and Blur Attack teams. Specifically to handle images with JPEG compression artifacts, CAPP_OB team used auto-endocer loss [36]. Dilated convolutions were adopted in many solutions to enlarge the receptive field. Also, the attention modules were widely employed.

4.2. Challenge Winners

The challenge winners are determined by the PSNR scores. In track 1, VIDAR team achieved the best restoration quality with their EDPN architecture. They also exhibited the best SSIM and LPIPS scores in track 2. The EDPN model is inspired from EDVR [84] and exploits the similarity information within the extracted features. In track 2, The Fat, The Thin and The Strong team showed the best PSNR score from their proposed HINet model. They propose Half Instance Normalization Block to design their model architecture.

4.3. Visual Comparison

We provide a visual comparison between the top-ranked solutions. Figure 1 shows the deblurred images from low-resolution input in Track 1. Figure 2 illustrates the images deblurred from JPEG-compressed input in Track 2.
5. Challenge Methods and Teams

5.1. VIDAR

VIDAR team proposed Enhanced Deep Pyramid Network (EDPN) [90] for blurry image restoration from multiple degradations. The overall structure of EDPN is shown in Figure 3, which is inspired by EDVR [84]. Specifically, they exploit the self- and cross-scale similarities in the degraded image with two pyramid-based modules, i.e., the pyramid progressive transfer (PPT) module and the pyramid self-attention (PSA) module. They first replicate the given blurry image K times (K = 4) and feed the replicated images as the input to EDPN, which aims to fully exploit the self-similarity contained in the degraded image. The features extracted from the multiple same images by a feature extractor consisting of 18 residual blocks are fed into the PPT module. The PPT module is designed to transfer the cross-scale similarity information from the same degraded image at the feature level with a pyramid structure, which performs the deformable convolution and generates attention masks to transfer the self-similarity information in a progressive manner. The following PSA module is designed to aggregate information across the above transferred features, which adopts the self- and spatial-attention mechanisms to fuse the multiple features. For the blurry image super-resolution task, the fused features are fed into a reconstruction module followed by an upsampling layer. For the blurry image deblocking task, the upsampling layer will not be necessary. The reconstruction module is composed of 120 multi-scale residual channel attention blocks [108]. Please refer to [90] for more details.

5.2. netai

netai team proposed Pixel-Guided Dual-Branch Attention Network (PDAN) for joint image deblurring and super-resolution. The dual-branch scheme of PDAN is similar to [107]. In PDAN, the feature extraction module uses residual spatial and channel attention (RSCA) module, inspired by [108]. The deblurring module is a residual encoder-decoder model to enlarge the receptive field, activated by LeakyReLU layers [56]. The shallow feature from the feature extraction module is fed into the reconstruction module to increase the spatial resolution. The upsampling is done by scale 4 through a convolutional layer to reconstruct the HR output image. netai team proposed an HPEM loss function for using a hard example mining strategy to focus on the difficult areas automatically. The whole model is jointly trained from scratch using L1 loss and then fine-tuned with the weighted sum of L1 loss and the HPEM loss. The overall architecture is shown in Figure 4. Please refer to [71] for more details.

5.3. NJUST-IMAG

NJUST-IMAG team developed an end-to-end network consisting of a deblurring module and a subsequent super-resolution module. A non-local residual network (NLRN) is proposed as the super-resolution module to better generate high-quality images. In the NLRN, the non-local residual group is adopted as the basic unit. The non-local residual group contains two sub-groups that each consist of a non-local block [85] and four RCABs [108]. The non-local architecture is effective at modeling the global information which is able to help remove the residual blur and further improve the super-resolution performance. Self-attention is
adapted to explore the relation between each image patch. Multi-head mechanism [82] is used to make the non-local block focus on more diverse global correlation.

The whole model is jointly trained starting from the pre-trained deblurring module. L1 loss and the image gradient loss are employed to train the model. The overall architecture is shown in Figure 5. More information can be found in [5].

5.4. SRC-B

SRC-B team proposed a Multi-Refinement Network (MRNet) for image deblurring. MRNet was originally developed for defocus deblurring on images from dual camera and applied to single image deblurring in this competition. MRNet is composed of 4 modules: feature extraction, fusion, reconstruction, and upsampling. The feature extraction module computes Siamese feature from the single input image. The features are concatenated in channel dimension and then fused by 1 × 1 convolution. Inspired by MMDM [51], Residual Block Module (RBM) is proposed. RBM adopts the same configuration as MMDM, consisting of 10 residual modules and a global residual connection. To avoid the increment in computational complexity, channel attention is not used. Similarly to FERM in [51], 5 RBM modules are used to form residual group module (RGM). Multi-scale RGM (MSRGM) is constructed from the RGMs by computing parallel features with encoder-decoder structure. Finally, the reconstruction module is composed of multiple MSRG modules, connected in series. On the idea that each block refines the features from the previous layer, every module has a global residual connection.

With the proposed architecture self-ensemble did not consistently increase PSNR. Thus, multi-model ensemble strategy was used to make final results. The overall architecture is shown in Figure 6.

5.5. Baidu

Baidu team proposes to improve MPRNet [101] and a RRDB-based model [86] and exploit the virtue of 2 models via ensemble. The MPRNet is enhanced by adding an upsampler in the 3rd stage and by introducing an iterative process in the SAM module. RRDB model was pretrained from DF2K dataset, combining DIV2K [2] and Flickr2K [77, 45] datasets as [86]. Each model was trained with L1, FFT, and MS-SSIM loss with large patches of size 320 × 320. For the enhanced MPRNet, 640 × 640 patches were used. The final output is generated from the ensemble of each model output that is an self-ensemble [78] result from 8 geometric transforms. The learning rate is initialized as 1 × 10^{-4} and halved at 20k, 30k, 35k iterations. Adam optimizer was used. The overall architecture is shown in Figure 7.

5.6. MMM

MMM team proposed a M3Net model using multi-stage, multi-patch, and multi-resolution

MMM team proposed a M3Net model using multi-stage, multi-patch, multi-resolution strategy. The model is divided into 3 levels and input of the each layer is a downsampled low-resolution input split in non-overlapping patches. For the lower two stages, encoder-decoder architecture with different depth is employed to extract features of multiple scales. The top stage does not have such encoder-decoder structure to preserve the spatial high-frequency information.
In the encoder-decoder structure, the features at the same resolution are aggregated by concatenation and convolution. They are progressively fused with the upper stage. CAB and ORB modules in [101] are applied at each stage to extract features. The skip connection between the encoders and decoders and the global skip connection with 4 times upsampling are introduced to enhance the image restoration quality. The overall architecture is shown in Figure 8.

5.7. Imagination

Figure 9: Imagination team (Track 1). Pyramid Deformable Convolution

Imagination team proposed a pyramid deformable convolution method. Borrowing the idea of PCD alignment module in EDVR [84], they used pyramid cascading DCN to further align individual image features. For the refined bicubic LR aligned features, an RCAN model [108] with 10 residual groups with 20 RCABs are applied. REDS120fps dataset [59] is used to synthesize additional training data. L2 loss is used at training. In the testing phase, 6 independent models with ×8 self-ensemble is used to obtain additional gains in PSNR. The overall architecture is shown in Figure 9.

5.8. Noah_CVlab

Figure 10: Noah_CVlab team (Track 1 & 2). Pre-trained Image Processing Transformer

Noah_CVlab team adopted an Image Processing Transformer (IPT) approach proposed in [10]. The IPT model consists of multi-head and multi-tail for different tasks and a shared transformer body including an encoder and a decoder. The input image is first converted to visual features and then divided into patches as visual words for subsequent processing. The resulting image with high visual quality is reconstructed by ensembling output patches.

In the pretraining phase, there are 6 heads and tails corresponding to the six image-to-image tasks including super-resolution with scale 2, 3, 4, denoising with noise level 30 and 50, and deraining. In the fine-tuning phase, the head and tail for ×4 super-resolution is chosen and the other heads and tails are dropped. Both the heads and tails are convolutional layers. The body consists of a 12-layer transformer encoder and a 12-layer transformer decoder.

The model is pretrained with ImageNet [14] dataset and fine-tuned with GOPRO [60] and REDS [59] datasets. ImageNet data is utilized for generating degraded images by downsampling, adding Gaussian noise, rain streaks. The overall architecture is shown in Figure 10.

5.9. TeamInception

Figure 11: TeamInception (Track 1 & 2). Multi-Stage Progressive Image Restoration

TeamInception presented MPRNet architecture introduced in [101]. MPRNet consists of three stages to progressively restore images. The first two stages are geased on encoder-decoder subnetworks that learn the broad contextual information due to the large receptive field. The last stage employs a subnetwork, ORSNet containing multiple ORB modules. Supervised attention module (SAM) is incorporated between the stages. Cross-scale feature fusion mechanism is introduced where the intermediate multi-scale contextual features of the earlier subnetworks help consolidating the intermediate features of the latter subnetwork. L1, MS-SSIM, VGG loss is used to train the model. The overall architecture is shown in Figure 11.

5.10. ZOCS_Team

ZOCS_Team used RDN [109] as a baseline and added token-based transformer to the building block, RDB. The
advantages of token-based transformer can be listed as follows: 1) similar patterns in an image are grouped 2) transformers use the non-local self-similarity based on image tokens 3) less computation cost is required compared with a non-local layer. The token-based transformer module is added after the last convolution layer of RDB. Then pyramid token visual transformer is added after the upsampling layer of RDN. The model is pretrained on DF2K dataset \([2, 45]\) is used and then fine-tuned on the REDS dataset \([59]\). The overall architecture is shown in Figure 12.

5.11. Mier

Mier team proposed a Big UNet based on MWCNN \([50]\) and RCAN \([108]\). They replaced the convolutional layers in MWCNN with the residual group in RCAN to enhance the reconstruction quality. In order to further expand the receptive field, they added a multi-scale dilated block (MDB) from DAVANet \([113]\). For track 1, bicubic upsampling is applied to the input to match the image resolution. \(\times 8\) self-ensemble is applied. The overall architecture is shown in Figure 13.

5.12. INFINITY

INFINITY team used EDSR \([45]\) to deblur images in Track 1. Self-ensemble was used with 8 geometric transforms. They tested WDSR \([97]\), RFDN \([48]\), DRN \([25]\), RCAN \([108]\), RCAN with pixel attention \([110]\) and self-calibrated convolutions \([49]\) and chose EDSR for better accuracy.

5.13. DMLAB

DMLAB team proposed Multi-scale Hierarchical Dense Residual Network (MS-HDRN). Hierarchical dense residual learning is proposed via multi-level dense connections and multi-level residual connections. To implement multi-level dense connection, \(1 \times 1\) convolution layers are inserted as the first and the last layers of MDCG and MDCB modules \([41]\), reducing the number of feature maps. Inspired by \([7]\), multi-scale feature extraction modules are used without reducing spatial resolution. To achieve the implementation principle, MDCB \([41]\) and Laplacian attention \([4]\) modules are used with modifications. The overall architecture is shown in Figure 14.

5.14. RTQSA-Lab

RTQSA-Lab team proposed Enhanced Attention Network for the competition track 1. They presented a new attention network consisting of a Global Attention Module (GAM) and a Local Attention Module (LAM) to model the dependencies between layers, channels, and positions. Specifically, the proposed GAM adaptively emphasizes hierarchical features by considering correlations among layers. Meanwhile, LAM learns the confidence at all positions of each channel, selectively capturing more informative features.

5.15. Yonsei-MCML

Yonsei-MCML team proposed an edge detection-based attention network for image deblurring. On top of BANet \([80]\) using dilated convolutions, attention module is added. The edge information is fed into the model by Sobel filter in the horizontal and the vertical directions. Following ESPCN \([70]\), sub-pixel convolution is employed with modification. The overall architecture is shown in Figure 15.
5.16. SCUT-ZS

SCUT-ZS team applied EDSR [45] in image deblurring task in participation to track 1.

5.17. Expasoft team

Expasoft team proposed a BowNet architecture, combining ESRGAN [86] and UNet [68]. The presented UNet consists of RRDB blocks [86] where the changes in the number of channels are made by $1 \times 1$ convolutions. Average pooling is used to reduce the scale of feature maps in UNet. The UNet and the ESRGAN body are applied in parallel and the extracted features are concatenated and fused in the next layers. The resolution is increased by a sequence of $3 \times 3$ convolutions and nearest-neighbor upsampling. The overall architecture is shown in Figure 16.

5.18. The Fat, The Thin and The Strong

The Fat, The Thin, and The Strong team proposes a two-stage feature completion network. Five feature representation for each stage is presented to effectively increase the receptive field. At each stage, a convolutional feature is extracted followed by a body architecture similar to UNet [68]. To replace the batch normalization, half-IN block is designed in the encoding stage. Half-IN block uses both the non-normalized and the normalized feature from instance normalization [81]. SAM block from MPRNet [101] is adopted to refine feature and interact with the input features of the second stage. 3 models were used for ensemble but not much PSNR boost were observed in the REDS validation set. The overall architecture is shown in Figure 17. Please refer to [11] for more details.

5.19. CAPP_OB

CAPP_OB team proposed a wide receptive field and channel attention network (WRCAN), an encoder-decoder architecture similar to UNet [68]. Dilated convolutions are used to increase the receptive field and the channel attention [108] considers the relation between the feature channels. They further optimize the model using with auto-encoder loss [36] to handle the JPEG compression artifacts. The overall architecture is shown in Figure 18. Please refer to [39] for more details.
5.20. DuLang

Figure 19: **DuLang team (Track 2)**. Multi-Scale Fusion Net

DuLang team proposes a multi-scale Fusion Net (MSFN) based on AFN [88] and MIRNet [100] to restore blurry images with JPEG artifacts. To expand the receptive field, dilated convolutions are added to Res-Block. Triple attention computes the attention weights by capturing cross-dimension interaction using a three-branch structure [58]. To train the proposed model, L1 loss, L1 loss between the Laplacian images, and the L1 loss between the Sobel-filtered images are used. The overall architecture is shown in Figure 19.

5.21. GiantPandaCV

Figure 20: **GiantPandaCV team (Track 2)**. A Simple Dilated Encoder-Decoder Network

GiantPandaCV team proposed a simple encoder-decoder structure model. Different from U-Net, the receptive field is enlarged by dilated convolution layers. They used SSIM and Charbonnier loss function to train the proposed model. The overall architecture is shown in Figure 20.

5.22. Maradona

Maradona team used a multi-scale residual network model [60] by extending the model depth to 182 layers in their participation in track 2. ×8 self-ensemble was used at test time.

5.23. LAB FUD

Figure 21: **LAB FUD team (Track 2)**. yuv-grid-net

LAB FUD team proposed a yuv-grid-net. The model converts the input sRGB image to YUV colorspace and concatenates a grid map in 8 × 8 and 16 × 16 size. Processed by a residual network, the output of the model is converted from the YUV colorspace to sRGB. The overall architecture is shown in Figure 21.

5.24. SYJ

SYJ team proposed a Multi-level Wavelet-ResNet. The proposed method performs discrete wavelet transforms in the neural network. Residual group modules [108] are used in the model with intermediate residual connections. The overall architecture is shown in Figure 22.

5.25. Dseny

Dseny team presented a multi-scale and multi-patch network to deblur images in real-time. They combined SRN [76] and DMPHN [102] to build their model architecture. The overall architecture is shown in Figure 23.

5.26. IPCV_IITM

IPCV_IITM team used multi-scale context block [96] in multi-patch hierarchical [102, 74] architecture. The overall architecture is shown in Figure 24.
5.27. Blur Attack

Blur Attack team proposed a model named EACD. The model extracts edge in addition to the convolutional feature. The features are processed by residual dense blocks $[109]$ and residual groups $[108]$. The overall architecture is shown in Figure 25.

Acknowledgments

We thank the NTIRE 2021 sponsors: HUAWEI Technologies Co. Ltd., Wright Brothers Institute, Facebook Reality Labs, MediaTek and ETH Zurich (Computer Vision Lab).

A. Teams and affiliations

NTIRE 2021 team

Title: NTIRE 2021 Challenge on Image Deblurring
Members: Seungjun Nah$^1$ (seungjun.nah@gmail.com), Sanghyun Son$^1$, Suyoung Lee$^1$, Radu Timofte$^2$, Kyoung Mu Lee$^1$
Affiliations: $^1$ Department of ECE, ASRI, SNU, Korea $^2$ Computer Science and Technology of China, China

VIDAR

Title: EDPN: Enhanced Deep Pyramid Network for Blurry Image Restoration
Members: Zhiwei Xiong (zwxiong@ustc.edu.cn), Ruikang Xu, Zeyu Xiao, Jie Huang, Yueyi Zhang
Affiliations: University of Science and Technology of China, China

The Fat, The Thin and The Strong

Title: HIINet: Half Instance Normalization Network for Image Restoration
Members: Liangyu Chen$^1$ (chenliangyu@megvii.com), Jie Zhang$^2$, Xin Lu$^1$, Xiaojie Chu$^3$, Chengpeng Chen$^1$
Affiliations: $^1$ Megvii $^2$ Fudan University, China $^3$ Peking University, China

netai

Title: Enhanced Multi-task Network for Blur Image Super-Resolution
Members: Si Xi (xisi789@126.com), Jia Wei
Affiliations: Netease Games AI Lab

NJUST-IMAG

Title: Learning A Cascaded Non-Local Residual Network for Super-resolving Blurry Images
Members: Haoran Bai$^1$ (baihaoran@njust.edu.cn), Songsheng Cheng$^1$, Hao Wei$^1$, Long Sun$^2$, Jinhui Tang$^1$, Jinshan Pan$^1$
Affiliations: $^1$ Nanjing University of Science and Technology, China $^2$ Guilin University of Electronic Technology, China

CAPP OB

Title: Wide Receptive Field and Channel Attention Network for Deblurring of JPEG Compressed Image
Members: Donghyeon Lee$^1$ (donghyeon1223@gmail.com), Chulhee Lee$^1$, Taesung Kim$^2$
Affiliations: $^1$ Samsung Electronics Co., Ltd. $^2$ Sunmoon University, Asan, Korea

SRC-B

Title: MRNet: Multi-Refinement Network for Image Deblurring
Members: Xiaobing Wang (x0106.wang@samsung.com), Dafeng Zhang
Affiliations:
Samsung Research China - Beijing

BAIDU
Title: Joint Super-Resolution and Deblurring Using Dual-Model Ensemble
Members: Zhihong Pan¹ (zhihongpan@baidu.com), Tianwei Lin², Wenhao Wu², Dongliang He³, Baopu Li¹, Boyun Li², Teng Xi², Gang Zhang², Jing tuo Liu², Junyu Han², Errui Ding³
Affiliations:
¹ Baidu Research USA
² Department of Computer Vision Technology, Baidu, Inc.

MMM
Title: M3Net: Multi-stage, Multi-patch and Multi-resolution for image deblurring
Members: Jie Zhang¹ (j.jiang19@fudan.edu.cn), Liangyu Chen², Chengpeng Chen², Xin Lu², Xiaojie Chu³
Affiliations:
¹ Fudan University, China
² Megvii
³ Peking University, China

Imagination
Title: Pyramid Deformable Convolution
Members: Guangping Tao¹ (tgpin@smail.nju.edu.cn), Wenqing Chu², Yun Cao², Donghao Luo², Ying Tai², Tong Lu¹, Chengjie Wang², Jilin Li², Feiyue Huang²
Affiliations:
¹ Nanjing University, China
² Tencent

Noah.CVlab
Title: Pre-Trained Image Processing Transformer
Members: Hatimg Chen (chenhanting@huawei.com), Shuaijun Chen, Tianyu Guo, Yunhe Wang
Affiliations:
Noah’s Ark Lab, Huawei Technologies Co., Ltd.

TeamInception
Title: Multi-Stage Progressive Image Restoration
Members: Syed Waqas Zamir (waqas.zamir@inceptioni.ai.org), Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Kahn, Ling Shao
Affiliations:
Inception Institute of Artificial Intelligence

ZOCS.Team
Title: PyVTRDN - Equip RDN with Pyramid token based transformer
Members: Yushen Zuo (zys19@mails.tsinghua.edu.cn), Yimin Ou, Yuanjun Chai, Lei Shi
Affiliations:
Tsinghua University, Beijing, China

Mier
Title: Big UNet for Image Restoration
Members: Shuai Liu¹ (18601200232@163.com), Lei Lei², Chaoyu Feng²
Affiliations:
¹ North China University of Technology ² Xiaomi

INFINITY
Title: Image Deblurring with Enhanced Deep Residual Network
Members: Kai Zeng (eezk@mail.scut.edu.cn), Yuying Yao, Xinran Liu
Affiliations:
South China University of Technology, China

DuLang
Title: Multi-Scale Fusion Net for Single Image Deblurring
Members: Zhizhou Zhang (220201874@seu.edu.cn), Huacheng Huang
Affiliations:
Lab of Image Science and Technology, Southeast University, China

GiantPandaCV
Title: A Simple Dilated Encoder-Decoder Network for Image Restoration
Members: Yunchen Zhang (cydiachen@cydiachen.tech), Mingchao Jiang, Wenbin Zou
Affiliations:
¹ China Design Group Co., Ltd. ² JOYY AI GROUP ³ Fujian Normal University, China

Maradona
Title: Multi-scale Network for Image Deblurring
Members: Si Miao (miaosi2018@gmail.com)
Affiliations:
Shanghai Advanced Research Institute, Chinese Academy of Sciences

LAB FUD
Title: yuv-grid-net
Members: Yangwoo Kim (tuzm24@gmail.com)
SYJ

Title: Multi-level Wavelet-Resnet in Residual learning  
Members: Yuejin Sun (yjsun97@mail.ustc.edu.cn)  
Affiliations: University of Science and Technology of China, China

Dseny

Title: Multi-scale and Multi-patch Network for High-Definition Image Deblurring in Real-time  
Members: Senyou Deng¹ (dengsenyou@iie.ac.cn), Wenqi Ren¹, Xiaochun Cao¹, Tao Wang²  
Affiliations: ¹ Institute of Information Engineering, Chinese Academy of Sciences, China ² Huawei Noah’s Ark Lab

IPCV_IITM

Title: Hierarchical Encoder-Decoder with Multi-scale Convolution for Image Deblurring  
Members: Maitreya Suin (maitreyasuin21@gmail.com), A. N. Rajagopalan  
Affiliations: Indian Institute of Technology Madras, India

DMLAB

Title: Multi-scale Hierarchical Dense Residual Network for Image Deblurring  
Members: Vinh Van Duong (duongvinh@skku.edu), Thuc Huu Nguyen, Jonghoon Yim and Byeungwoo Jeon  
Affiliations: Department of ECE, Sungkyunkwan University, Korea

RTQSA-Lab

Title: Enhanced Attention Network for Single Image Deblurring  
Members: Ru Li¹ (leeru564636@163.com), Junwei Xie²  
Affiliations: ¹ Fuzhou University, China ² Imperial Vision Co. Ltd.

Yonsei-MCML

Title: Edge Attention Network for Image Deblurring and Super-Resolution  
Members: Jong-Wook Han (johnhan00@yonsei.ac.kr), Jun-Ho Choi, Jun-Hyuk Kim, Jong-Seok Lee  
Affiliations: School of Integrated Technology, Yonsei University

SCUT-ZS

Title: EDSR for NTIRE 2021 Image Deblurring Challenge  
Members: Jiaxin Zhang (msjxzhang@mail.scut.edu.cn), Fan Peng  
Affiliations: South China University of Technology, China

Expasoft team

Title: BowNet  
Members: David Svitov (d.svitov@expasoft.tech), Dmitry Pakulich  
Affiliations: ¹ Expasoft LLC ² Institute of Automation and Electrometry of the SB RAS

Blur Attack

Title: EACD: Deblurring Network Using Edge Module, ASPP Channel Attention and Dual Network  
Members: Jaeyeob Kim (athurk94111@gmail.com), Jechang Jeong  
Affiliations: Image Communication & Signal Processing Laboratory, Hanyang University, Korea
References


Instance normalization: The missing ingredient for fast pretraining.


