NTIRE 2020 Challenge on Perceptual Extreme Super-Resolution: Methods and Results

Kai Zhang Shuhang Gu Radu Timofte Taizhang Shang Qiuju Dai
Shengchen Zhu Tong Yang Yandong Guo Younghyun Jo Sejong Yang
Seon Joo Kim Lin Zha Jiande Jiang Xinbo Gao Wen Lu Jing Liu
Kwangjin Yoon Taegyun Jeon Kazutoshi Akita Takeru Ooba Norimichi Ukita
Zhipeng Luo Yuehan Yao Zhenyu Xu Dongliang He Wenhao Wu Yukang Ding
Chao Li Fu Li Shilei Wen Jianwei Li Fuzhi Yang Huan Yang Jianlong Fu
Byung-Hoon Kim JaeHyun Baek Jong Chul Ye Yuchen Fan Thomas S. Huang
Junyeop Lee Bokyeung Lee Jungki Min Gwantae Kim Kanghyu Lee
Jaihyun Park Mykola Mykhailych Haoyu Zhong Yukai Shi Xiaojun Yang
Zhijing Yang Liang Lin Tongtong Zhao Jinjia Peng Huibing Wang Zhi Jin
Jiahao Wu Yifu Chen Chenming Shang Huanrong Zhang Jeongki Min
Hrishikesh P S Densen Puthussery Jiji C V

Abstract

This paper reviews the NTIRE 2020 challenge on perceptual extreme super-resolution with focus on proposed solutions and results. The challenge task was to super-resolve an input image with a magnification factor $\times 16$ based on a set of prior examples of low and corresponding high resolution images. The goal is to obtain a network design capable to produce high resolution results with the best perceptual quality and similar to the ground truth. The track had 280 registered participants, and 19 teams submitted the final results. They gauge the state-of-the-art in single image super-resolution.

1. Introduction

Recent years have witnessed tremendous success of using deep neural networks (DNNs) to generate a high-resolution (HR) image from a low-dimensional input [5, 8, 12, 31, 42, 51]. On the one hand, DNNs-based single image super-resolution (SR) for bicubic degradation is continuously showing improvements in terms of PSNR and perceptual quality [5, 9, 12, 16, 25, 40, 41, 42]. In particular, several fundamental conclusions have been drawn: (i) DNNs based SR with pixel-wise loss (such as L1 loss and L2 loss) tends to produce oversmoothed output for a large scale factor due to the pixel-wise average problem [22]; (ii) The perceptual quality of super-resolved image could be improved by using VGG perceptual loss and generative adversarial (GAN) loss [11, 22, 45]; (iii) There is a trade-off between reconstruction accuracy and perceptual quality, which means no DNNs-based method can achieve its best PSNR and best perceptual quality at the same time [5]. While perceptual SR for bicubic degradation at a moderate scale factor (*e.g.*, $\times 4$) has achieved significant progress [22, 45, 53, 54, 57], the case with an extremely large scale factor has received little attention [7, 12]. On the other hand, realistic HR image synthesis from a latent low-dimensional vector based on GAN has shown great success for natural image [6] and face image [21]. However, how to effectively generate a perceptually pleasant HR image from a low-resolution (LR) image with a very large scale factor remains an open problem.

Jointly with NTIRE 2020 workshop we have an NTIRE challenge on perceptual extreme super-resolution, that is, the task of super-resolving an LR image to a perceptually pleasant HR image with a magnification factor $\times 16$. Although AIM 2019 extreme SR challenge [12] has considered the fidelity track and perceptual track, it has been concluded that: (i) the PSNR-orientated methods consistently give rise to oversmoothed results; (ii) there still remains a large room for perceptual quality improvement. As a result, this challenge only has one track which aims to seek effective solutions for perceptual extreme SR.

K. Zhang (kai.zhang@vision.ee.ethz.ch, ETH Zurich), S. Gu, and R. Timofte are the challenge organizers, while the other authors participated in the challenge. Appendix A contains the authors' teams and affiliations.
NTIRE webpage:
https://data.vision.ee.ethz.ch/cvl/ntire20/

2. NTIRE 2020 Challenge

This challenge is one of the NTIRE 2020 associated challenges on: deblurring [33], nonhomogeneous dehazing [3], perceptual extreme super-resolution [52], video quality mapping [10], real image denoising [1], real-world super-resolution [30], spectral reconstruction from RGB image [4] and demoireing [49].

The objectives of the NTIRE 2020 challenge on perceptual extreme super-resolution challenge are: (i) to advance researches on perceptual SR at an extremely large scale factor; (ii) to compare the effectiveness of different methods and (iii) to offer an opportunity for academic and industrial attendees to interact and explore collaborations.

2.1. DIV8K Dataset [13]

Following [13], the DIV8K dataset which contains 1,700 DIVerse 8K resolution RGB images is employed in this challenge. The HR DIV8K is divided into 1,500 training images, 100 validation images and 100 testing images. The corresponding LR images in this challenge is obtained via default setting (bicubic interpolation) of Matlab function `imresize` with scale factor 16. The testing HR images are completely hidden from the participants during the whole challenge. In order to get access to the data and submit the testing HR results, registration on Codalab (https://competitions.codalab.org/) is required.

2.2. Track and Competition

Track This challenge has only one track. The aim is to obtain a network design capable to produce high resolution results with the best perceptual quality and similar to the ground truth.

Challenge phases *(1) Development phase:* the participants got the 1,500 HR training images and 100 LR validation images of the DIV8K dataset; the participants got the LR training images via Matlab's `imresize` function. Due to the storage constraints, the participants uploaded the center 1,000×1,000 HR validation results to an online validation server to get immediate feedback. During this phase, the provided feedback consisted from Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM) [46] results. However, since both PSNR and SSIM are not suitable for perceptual ranking, the validation results are only used to test whether the cropped images are correctly uploaded. *(2) Testing phase:* the participants got 100 LR testing images; the participants submitted their center 1,000×1,000 results of the super-resolved HR images to Codalab and emailed the code and factsheet to the organizers; the participants got the final results after the end of the challenge.

Evaluation protocol Apart from PSNR and SSIM, the quantitative measures also includes LPIPS (Learned Perceptual Image Patch Similarity) [55] and no-reference PI (Perceptual Index) [5, 32] which have been acknowledged as useful perceptual metrics. The evaluation is performed on the center 1,000×1,000 HR results for convenience and consistency with the reported results on the challenge servers. The final ranking employed also a user study. In order to have a thorough evaluation, the self-reported number of parameters and running time per testing image are also reported.

3. Challenge Results

From 280 registered participants, 19 teams entered in the final phase and submitted results, codes, and factsheets. Table 1 reports the final test results and rankings of the challenge. Note that the methods trained with GAN loss are grouped together, and only the best six methods are ranked by user study. The results of winner teams in AIM 2019 extreme SR challenge are also reported for comparison. Figures 1 and 2 show the visual results and associated PSNR, SSIM, LPIPS and PI values of different methods.

From Table 1 and Figures 1 and 2, we can have the following observations. First, the OPPO-Research team is the first place winner of this challenge, while CIPLAB and HiImageTeam win the second place and third place, respectively. Second, among the top-6 methods, ECNU achieves a good trade-off between reconstruction accuracy and perceptual quality, while HiImageTeam shows the best trade-off between number of parameters and inference time. Third, DeepBlueAI achieves the best PSNR performance, however, it fails to generate results with competitive perceptual quality. Fourth, LPIPS and PI are relatively reliable perceptual measures in comparison with PSNR and SSIM for perceptual extreme SR. Fifth, it is easy to distinguish the ground-truth HR images from the super-resolved HR images.

Architectures, losses and main ideas All the proposed methods utilize deep neural networks for perceptual extreme SR. Overall, there are two import factors to improve the perceptual quality of super-resolved images, *i.e.*, network architecture and loss function. For the network architecture, several teams, such as CIPLAB, APTX4869 and MSMers, proposed to extend existing state-of-the-art SR methods with a progressive upscaling strategy. Several other teams, such as OPPO-Research, SIA and DeepBlueAI, achieved a scale factor of 16 by directly modifying the upscaling layer. For the loss function, most of the teams adopted either L1 loss or the same loss (*i.e.*, a combination of L1 loss, VGG perceptual loss [19, 39] and relativistic GAN loss [20]) proposed in ESRGAN [45] as their final loss. In particular, CIPLAB replaced the VGG perceptual loss with LPIPS loss, and both CIPLAB and TTI adopted an U-Net [35]-like discriminator [36] for better local and global perceptual quality enhancement.

PSNR/SSIM/LPIPS (a) HR | 20.42/0.3270/0.426 (b) OPPO-Research | 20.31/0.3544/0.441 (c) CIPLAB | 20.90/0.3980/0.561 (d) HiImageTeam | 22.14/0.4904/0.657 (e) ECNU

20.92/0.4562/0.602 (f) SIA | 17.98/0.2875/0.505 (g) TTI | 22.26/0.5965/0.714 (h) DeepBlueAI | 22.20/0.4947/0.712 (i) APTX4869 | 21.88/0.4826/0.740 (j) CNDP-Lab

Figure 1. SR results (600×600 crop) by different methods.

PSNR/SSIM/FPIPS (a) HR | 19.98/0.4441/0.367 (b) OPPO-Research | 19.43/0.4273/0.370 (c) CIPLAB | 20.30/0.4990/0.410 (d) HiImageTeam | 22.14/0.5753/0.547 (e) ECNU

20.04/0.5328/0.486 (f) SIA | 15.56/0.3910/0.421 (g) TTI | 22.27/0.5812/0.616 (h) DeepBlueAI | 22.27/0.5796/0.626 (i) APTX4869 | 22.15/0.5769/0.621 (j) CNDP-Lab

Figure 2. SR results (500×500 crop) by different methods.

Ensembles Most of the teams adopted commonly-used model-ensemble or self-ensemble [43] to enhance the performance.

Train data Most of the teams only used the provided DIV8K dataset [13] for training. OPPO-Research further adopted DIV2K [2], Flickr2K [40] and OST [44] datasets, while APTX4869 and SuperT used DIV2K [2] as additional training data.

Conclusions From the above analysis of different solutions, we can have several conclusions. (i) The proposed methods improve the state-of-the-art for extreme SR. On one hand, compared to the best method proposed by NUAA-404 in AIM 2019 fidelity extreme SR challenge [12], DeepBlueAI achieved an average PSNR gain of 0.07dB. On the other hand, compared to the best method proposed by BOE-IOT-AIBD in AIM 2019 perceptual extreme SR challenge [12], OPPO-Research, CIPLAB and HiImageTeam produced perceptually better results and an

Table 1. Results of NTIRE 2020 perceptual extreme SR challenge. The PSNR, SSIM [46], LPIPS [55] and PI [5] are calculated on the center 1,000×1,000 subimages of the DIV8K test images.

Team	Author	PSNR ↑	SSIM ↑	LPIPS ↓	PI ↓	#Params [M]	Time [s]	Platf.	Ens.	GPU	Extra data	Loss
Methods optimized with GAN losses												
OPPO-Research	sss	$23.38_{(12)}$	$0.5504_{(15)}$	$0.348_{(1)}$	$3.97_{(2)}$	20.5	8.1	PyTorch	Model	V100	DF2K, OST	L1, VGG-P, GAN
CIPLAB	heyday097	$22.77_{(15)}$	$0.5251_{(16)}$	$0.352_{(2)}$	$3.76_{(1)}$	33.0	3.0	PyTorch	None	Xp	None	Huber, LPIPS, FM, GAN
HiImageTeam	HiImageTeam	$23.53_{(11)}$	$0.5624_{(13)}$	$0.368_{(3)}$	$4.38_{(4)}$	4.0	1.0	PyTorch	None	RTX	None	L1, VGG-P, GAN
Winner AIM19 [12]	*BOE-IOT-AIBD*	24.52	0.5800	0.418	6.28	-	47.1	PyTorch	None	Titan X	None	L1, VGG-P, GAN
ECNU	lj1995	$25.56_{(4)}$	$0.6336_{(6)}$	$0.497_{(6)}$	$8.10_{(8)}$	57.9	26.0	PyTorch	Self	1080Ti	None	L1, VGG-P, GAN
SIA	yoon28	$22.86_{(14)}$	$0.5896_{(11)}$	$0.434_{(5)}$	$5.81_{(6)}$	16.0	360.0	PyTorch	Self	CPU	None	L1, VGG-P, GAN
TTI	iim-nike	$19.16_{(17)}$	$0.4993_{(17)}$	$0.377_{(4)}$	$3.99_{(3)}$	26.5	19.5	PyTorch	None	V100	None	L1, VGG-P, GAN
sysu-AIR	Zhi_Jin_SYSU	$23.94_{(10)}$	$0.5545_{(14)}$	$0.510_{(7)}$	$4.99_{(5)}$	2.1	4.3	PyTorch	None	2080Ti	None	L1, TV, FS, VGG-P, GAN
CET_CVLab	hrishikeshps	$19.68_{(16)}$	$0.4290_{(18)}$	$0.705_{(19)}$	$7.46_{(7)}$	223.8	440.0	TensorFlow	None	CPU	None	L1, VGG-P, GAN
From here down are mostly L1/L2 optimized methods												
DeepBlueAI	DeepBlueAI	$25.70_{(1)}$	$0.6390_{(2)}$	$0.555_{(9)}$	$9.15_{(9)}$	63.5	0.8	PyTorch	Model	V100	None	L1; L2
Winner AIM19 [12]	*NUAA-404*	25.63	0.6394	0.554	9.21	-	30.0	PyTorch	Self	2080Ti	-	L1
APTX4869	APTX4869	$25.62_{(2)}$	$0.6393_{(1)}$	$0.558_{(11)}$	$9.25_{(11)}$	20.8	40.0	PyTorch	Self	V100	DIV2K	L2, VGG-P, Style, GAN; L1
CNDP-Lab	albertron	$25.58_{(3)}$	$0.6367_{(3)}$	$0.556_{(10)}$	$9.36_{(14)}$	322.8	350.0	PyTorch	Self	RTX	None	L1
MSMers	huayan	$25.56_{(4)}$	$0.6366_{(4)}$	$0.564_{(12)}$	$9.40_{(16)}$	16.0	16.0	PyTorch	Self	V100	None	L1
kaws	egyptdj	$25.43_{(5)}$	$0.6339_{(5)}$	$0.568_{(13)}$	$9.37_{(15)}$	126.2	9.76	TensorFlow	Self	V100	None	L1
UIUC-IFP	fyc0624	$25.28_{(7)}$	$0.6339_{(5)}$	$0.553_{(8)}$	$9.28_{(12)}$	100.6	38.0	PyTorch	Self	1080Ti	None	L1
KU_ISPLB	givenjiang	$25.33_{(6)}$	$0.6299_{(7)}$	$0.582_{(14)}$	$9.44_{(17)}$	1.9	70.0	Pytorch	Self	RTX	None	L1
MsSrModel	nickolay	$25.23_{(8)}$	$0.6259_{(8)}$	$0.586_{(15)}$	$9.29_{(13)}$	9.0	1.3	TensorFlow	Self	V100	None	L2, VGG-P
MoonCloud	pigfather315	$25.17_{(9)}$	$0.6250_{(9)}$	$0.587_{(16)}$	$9.23_{(10)}$	4.9	3.2	PyTorch	None	V100	None	L1
SuperT	tongtong	$23.94_{(10)}$	$0.6060_{(10)}$	$0.630_{(17)}$	$9.40_{(16)}$	0.4	0.6	Tensorflow	Self	V100	DIV2K	L1
KU_ISPL_A	jkm_ispl	$23.05_{(13)}$	$0.5787_{(12)}$	$0.667_{(18)}$	$11.4_{(18)}$	0.3	30.0	PyTorch	None	2080Ti	None	L1
Baseline	*Bicubic*	24.22	0.6017	0.683	11.1							

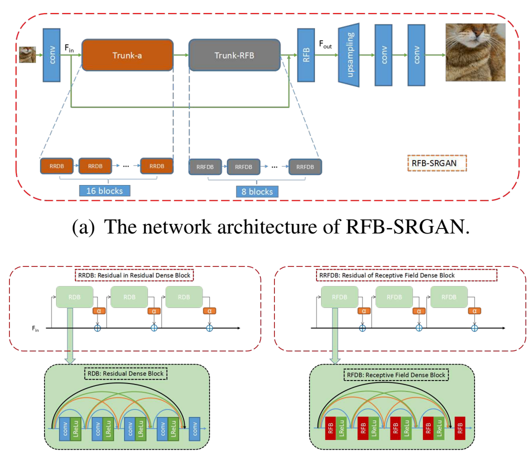
improved LPIPS value. (ii) The perceptual extreme SR is far from being solved. Advanced training strategies, network architectures and perceptual losses for a good trade-off between reconstruction accuracy and perceptual quality require further study. (iii) Due to the high ill-poseness of ×16 SR, DIV8K might not be sufficient to capture the diversity of natural images.

4. Challenge Methods and Teams

OPPO-Research

OPPO-Research proposed a novel Super-Resolution GAN [37], namely RFB-SRGAN, based on ESRGAN [45]. As shown in Figure 3, RFB-SRGAN consists of 5 parts, including the feature extraction module, the Trunk-a module, the Trunk-RFB module, the features up-sampling module, and the final convolution reconstruction module. Specifically, the feature extraction module is composed of a convolution layer. The Trunk-a module consists of 16 Residual in Residual Dense Blocks (RRDBs). The Trunk-RFB module is stacked of 8 Residual of Receptive Field Dense Blocks (RRFDBs), and each RRFDB contains 5 Receptive Fields Blocks (RFBs) [28]. The feature up-sampling module uses sub-pixel convolution [38] and nearest neighborhood interpolation which can greatly reduce the time cost while maintaining satisfactory performance. The final convolution reconstruction module consists of two layers of convolutions.

The training process is divided into two stages. In the first stage, a PSNR-oriented model was trained with L1 loss. The learning rate is initialized as 2×10^{-4} and decayed by a factor of 2 every 2×10^{5} steps. In the second stage, the generator was initialized by the pre-trained PSNR-oriented model and further trained with the loss function in ESRGAN [45]. The training mini-batch size is set to 16. The resolution of the cropped HR images is 512×512. The performance of models at different iteration stages are scored, then the top-scored models are fused to obtain the final model.



(a) The network architecture of RFB-SRGAN.

(b) RRDB block. (c) RRFDB block.

Figure 3. OPPO-Research's network architecture.

CIPLAB

CIPLAB proposed to use GAN [11] with LPIPS [55] loss for perceptual extreme SR [18] instead of using GAN with VGG perceptual loss [19]. Generally, the loss functions for perceptual SR is the adversarial loss [11] with the VGG perceptual loss [19]. Such a loss combination has worked well for ×4 SR, however, it is found that it is not work well for

×16 SR due to highly hallucinated noise and less precise details. Because VGG network is trained for image classification, it may not the best choice for the SR task. On the other hand, the learned perceptual similarity (LPIPS) [55] is trained with a dataset of human perceptual similarity judgments, thus it is expected to be a more proper choice for perceptual SR. For this reason, LPIPS is adopted instead of the VGG perceptual loss. In addition, the discriminator's feature matching loss helps to increase the quality of the results, and Huber loss prevents color permutation. The proposed generator, as shown in Figure 4, consists of two ESRGAN generators for ×16 SR. For the discriminator, an U-Net network as in [36] is adopted to judge real and fake for the compressed space from the encoder head and every pixel from the decoder head. Such a discriminator allows to provide detailed per-pixel feedback to the generator while maintaining the global context. It is empirically found that the discriminator can recover more details than normal encoder structure discriminator.

In the training, the HR patch size is set to 384×384, and the corresponding LR patch size is 24×24. For both generator and discriminator, Adam optimizer with learning rate 1×10^{-5} is adopted. The generator is first trained with L2 loss and mini-batch size 3 for 50K iterations. Then the model is trained with the proposed new combination of different loss functions and mini-batch size 2 for about 60K iterations.

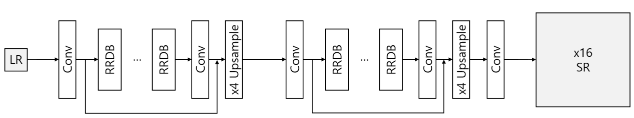


Figure 4. CIPLAB's generator network.

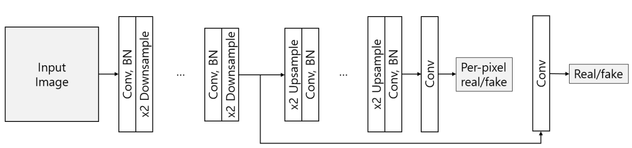


Figure 5. CIPLAB's discriminator network.

HiImageTeam

HiImageTeam proposed Cascade SR-GAN (CSRGAN) for perceptual extreme SR. As shown in Figure 6, CSRGAN achieves an upscaling of ×16 via four successive ×2 subnetworks (CSRB). In order to improve the performance, a novel residual dense channel attention block (see Figure 7) is proposed. Final CSRGAN uses VGG perceptual loss and GAN loss to enhance the perceptual quality of super-resolved images.

ECNU

ECNU proposed a Two Stages Super Resolution Generative Adversarial Network (TS-SRGAN). The network,

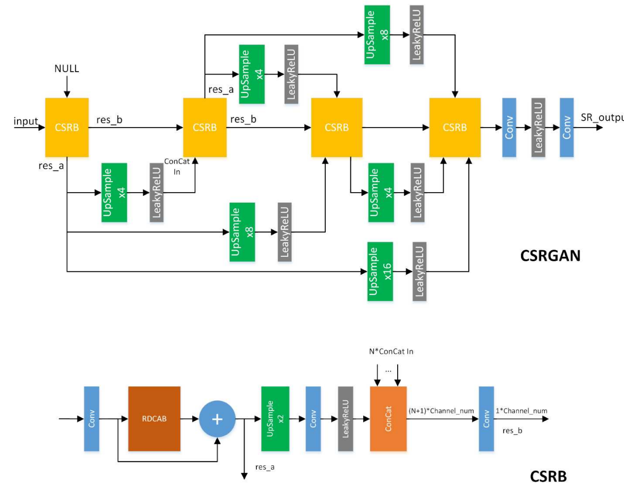


Figure 6. HiImageTeam's network architecture.

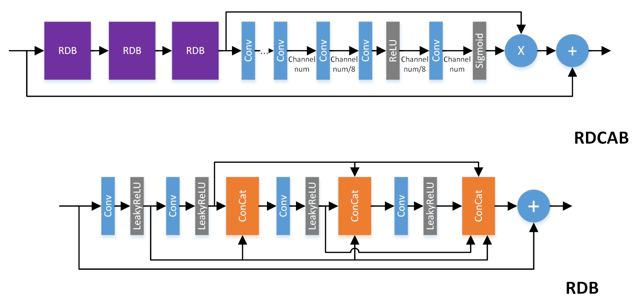


Figure 7. HiImageTeam's Residual Dense Channel Attention Block.

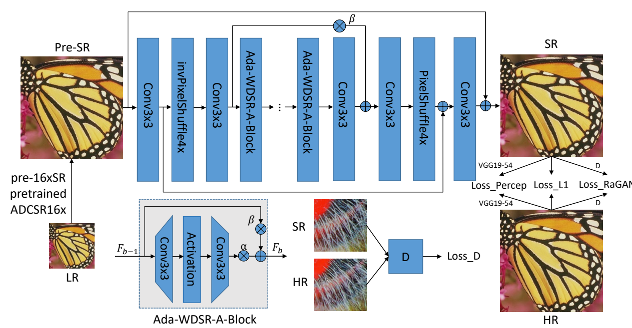


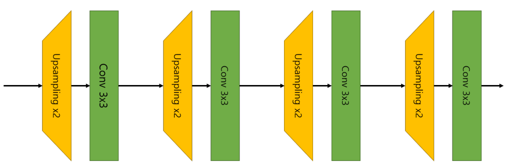
Figure 8. ECNU's network architecture.

as shown in Figure 8, consists of the pre-Super-Resolution (pre-SR) sub-net for the first stage and a main net for the second stage. Pretrained ADCSR16× [47] is used for the pre-SR sub-net to super-resolve the LR image to pre-SR image with the same resolution as ground truth. The pre-SR sub-net is frozen during the training phase of TS-SRGAN. The main sub-net is composed of a head 3×3 convolutional layer, a de-sub-pixel convolutional layer, a non-linear feature mapping module, an upsampling skip connection, a sub-pixel convolutional layer, a global skip connection and a tail 3×3 convolutional layer. The Adaptive WDSR-A-

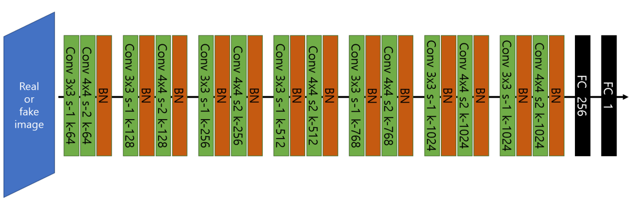
Block is modified from WDSR-A-Block [48] by adding learnable weight for body scaling and learnable weight for residual scaling. The loss function and the training processes are the same as ESRGAN [45].

SIA

SIA adopted ESRGAN [45] to extremely super-resolve an input image with a magnification factor of 16. The architecture of the generator is almost identical to that of ESRGAN except that the upscaling block of the generator consists of four upsampling layers, each of which doubles the feature size. The discriminator has seven downsampling blocks, each of which consists of two convolutional layers and batch-normalization (BN) layers [17] between them, while the first downsampling block has one BN layer. At the tail of the discriminator, two fully connected (FC) layers with size 256 and 1 is applied. See Figure 9 for the details. The hyper-parameters, such as optimizer, learning rate and coefficients for loss terms, are the same as ESRGAN.



(a) Upsampling block of generator



(b) Discriminator

Figure 9. SIA's network architecture.

TTI

Inspired by deep back-projection network [14] which introduces iterative up-down projection units for mutual relation between LR and HR feature maps, TTI proposed recurrent progressive perceptual DBPN where each up- and down-projection unit has one $\times 4$ and $\times \frac{1}{4}$ scaling layers and performs twice to expand LR images by a scale factor of $\times 16$ (see Figure 10). Such a network design not only reduces the number of model parameters but eases the training. For the discriminator, an UNet-like adversarial network (see Figure 11) inspired by [36] was used as it can capture both global features (*e.g.*, geometric or structural patterns) and local features (*e.g.*, texture patterns).

DeepBlueAI

DeepBlueAI proposed bag of tricks for perceptual extreme SR. Based on RCAN [57], various experiments were

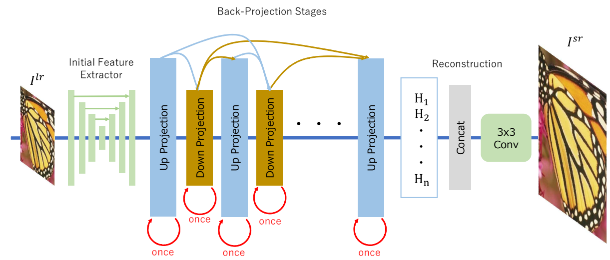


Figure 10. TTI's network architecture.

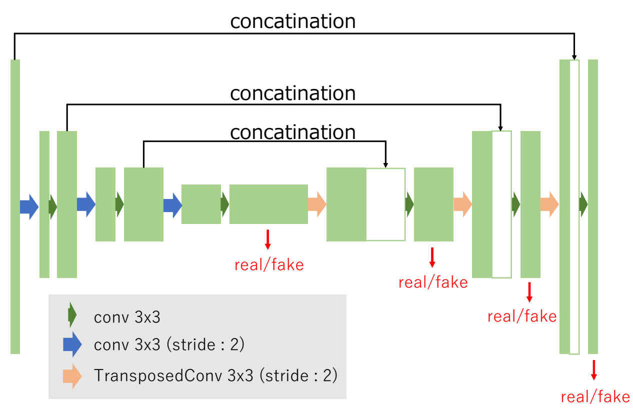


Figure 11. TTI's UNet-like discriminator.

conducted to explore how to improve PSNR and SSIM. According to the experiments, solutions based RCAN yielded best performance. In the final model, the number of residual groups is 10 and the number of channel in each layer is 128, the number of residual channel attention blocks in each residual group is 20. To obtain the $\times 16$ model, a $\times 4$ model is first trained from scratch as a pre-trained network after its convergence.

In the training, 1,500 training images with random horizontal flip and rotation are used. In each training batch, 16 LR patches with the size of 48×48 are extracted as inputs. The model is trained by ADAM optimizer with an initial leaning rate 1.0×10^{-4}. The learning rate utilize a cosine annealing schedule [29] with total 4.0×10^5 iterations and restarts every 1.0×10^5 iterations. Following [34], a $\times 16$ model with L1 loss is first trained and then is fine-tuned with L2 loss. To enhance the PSNR and SSIM, both self-ensemble and model-ensemble are utilized.

APTX4869

APTX4869 proposed progressive super-resolving and refining to tackle the perceptual extreme SR problem. The proposed network, as shown in Figure 12, is decomposed into two cascaded $4\times$ super-resolution ones with three training stages. In the first stage, the $\times 4$ DBPN [14] upsampling sub-module is trained to reconstruct an intermediate artifacts-free result. In the second stage, another $\times 4$ upsampling sub-module is then trained with generative adversarial settings to better recover details. In the third stage, the overall architecture is end-to-end refined with L1 loss to eliminate annoying artifacts.

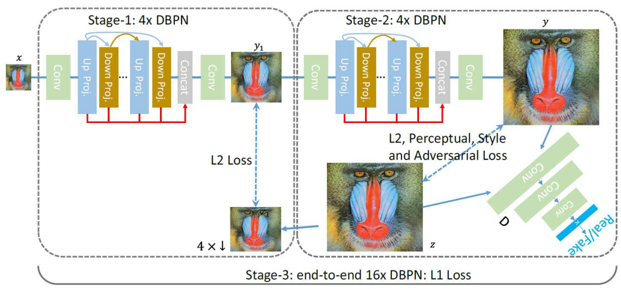


Figure 12. APTX4869's network architecture.

CNDP-Lab

CNDP-Lab proposed a cascaded U-Net with channel attention for extreme SR. The network, as shown in Figure 13, is developed from the traditional U-Net [35] by embedding channel attention [57] into the up-sampling process. Before concatenating the two feature maps of x_1 and x_2, the features are optimized by channel attention blocks, as shown in Figure 13(b). The down-sampling process is implemented by a convolutional layer with a stride of two, and the up-sampling process adopts the pixel shuffle method. After each down-sampling or up-sampling layer, a set of residual channel attention blocks are added to learn features at each scale. To up-scale the input low-resolution image with a factor of 16, two U-Net models are trained and each of the model up-scales the input image with a factor of 4. Because the input and output image of the U-Net have the same resolution, the input image is first up-scaled with a factor of 4 by bicubic interpolation.

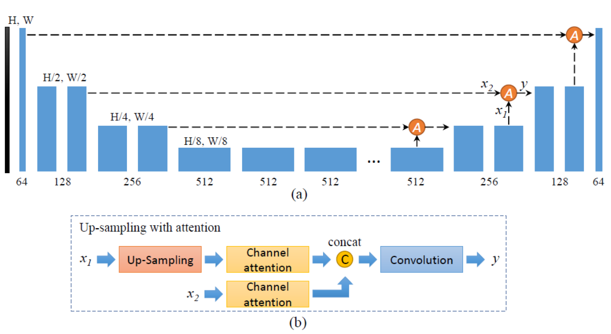


Figure 13. CNDP-Lab's U-Net with channel attention network architecture.

MSMers

The MSMers team proposed PURCAN which adopts progressive upsampling architecture and progressive training strategy for extreme SR. PURCAN takes RCAN [57] without the final upsampler as the backbone. The feature maps are progressively upsampled ($\times 1$ to $\times 4$ to $\times 16$) using pixel shuffle. The RCAN backbone is applied into $1\times$ scale while four residual channel attention blocks are applied into the $\times 4$ scale to enhance the performance. During training, the backbone part is first trained using $\times 4$ dataset which is constructed from DIV8K HR training images. Then the remaining part of the model is added to train on $\times 16$ DIV8K training images.

kaws

The kaws team proposed wavelet pyramid generation based high-frequency recovery for extreme SR. In the first stage, wavelet pyramid is generated from low-resolution (LR) image to upscale the image with explicitly recovered high-frequency details. The upscaled image is then refined from upscaled image to follow the ground-truth HR image in the second stage (see Figure 14(A)). To generate the wavelet detail coefficients from the LR image, they first concatenate outputs from convolving the LR image and sobel edge filters of four directions (vertical, horizontal, two-way diagonal). The concatenated data is input to the wavelet detail generator to output the LH, HL, HH wavelet detail coefficients (see Figure 14(B)). The proposed method is expected to have two advantages respect to image quality and generalizability. In terms of image quality, sharp edges can be recovered by upscaling the image with explicitly generated wavelet detail coefficients, which represent the local high-frequency information of each pixels. For generalizability, refining the upscaled image can be done with any state-of-the-art super-resolution (SR) model and is expected to improve its performance. The wavelet detail generator and the refine module are based on the modified U-Net [35] and the EDSR [27], respectively.

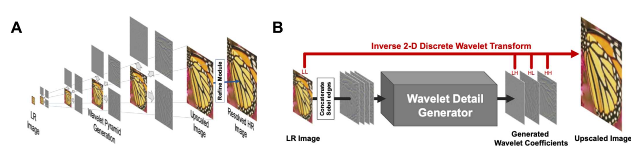


Figure 14. The network architecture of kaws team.

UIUC-IFP

Inspired by progressive multi-scale modelling, UIUC-IFP extended WDSR [48] from single-scale to multi-scale. The proposed progressive WDSR consists of 4 stages for $\times 16$ image SR. Starting from original feature spatial size, features are enlarge 2 times spatially and width are reduced 2 times at the end of each stage.

KU_ISPLB

KU_ISPLB proposed feedback recurrent neural network (FBRNN) for extreme SR [23]. As shown in Figure 15, the LR image is combined with improved LR every step. Inspired by GMFN [24] and DPBN [14], improved LR is given by upsampling network and down-projection network. The the image back projection network is performed 4 times after the each super resolution network. After yielding better LR image from the HR images in the process, the improved LR is combined with the original LR to prevent bias or over-fitting problems.

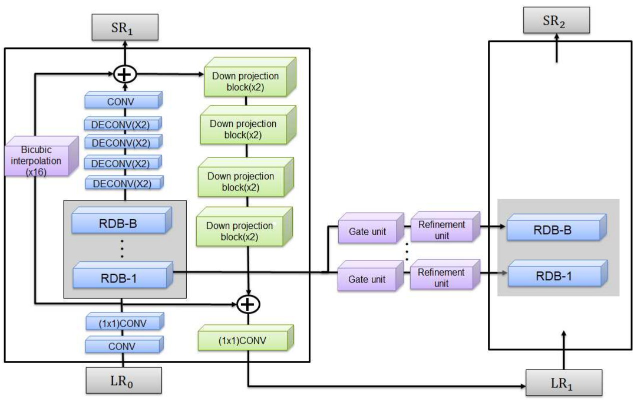


Figure 15. KU_ISPLB's network architecture.

MsSrModel

MsSrModel proposed multi-scale SR model (MsSrModel) (see Fig. 16) for extreme SR. The main idea is to have 5 different optimizations simultaneously for 5 different resolutions. The model operates on an LR image with pixel shuffle operation [38] at the end of each optimization. L2 loss and VGG perceptual loss are used to optimize the model.

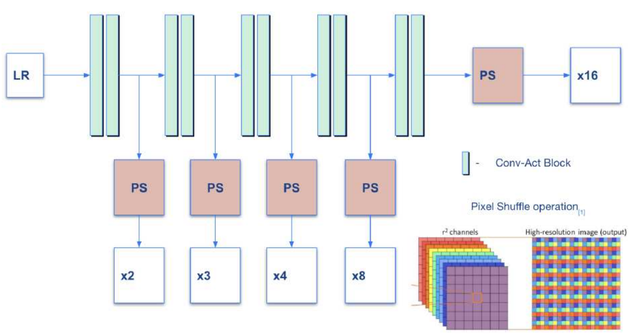


Figure 16. MsSrModel's network architecture.

MoonCloud

MoonCloud proposed Multi-scale ResNet for perceptual extreme SR. The network uses 16 residual blocks for feature extraction, then adopts a sub-pixel layer to construct the HR features, and finally incorporates a convolution layer with size of $8\times3\times3\times3$ to produce the final output. Based on SRResNet [22], a multi-scale strategy is used to learn rich feature for image restoration. In particular, multiple upsampling layers are used to implement multi-scale image SR.

SuperT

SuperT proposed fast and balanced Laplacian pyramid networks for progressive image super-resolution. The network, as shown in Figure 17, takes LR images as input and gradually predicts sub-band residuals from coarse to fine. At each level, the feature maps are first extracted to reconstruct a higher level image by using a lightweight upsampling module (LUM) with relatively sparse connections. Finally, convolutional layers are adopted to predict subband residuals. The prediction residuals at each level are used to efficiently reconstruct HR images through upsampling and addition operations.

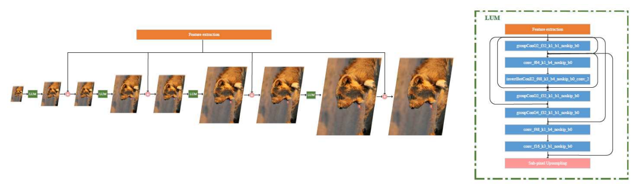


Figure 17. SuperT's network architecture.

sysu-AIR

The sysu-AIR team proposed a fast feedback network for large scale image super-resolution. Inspired by SRFBN [26] and IMDN [15], the proposed Fast-SRFBN is still reserved the RNN structure but with a information multi-distillation module (IMDM), which can benefit image SR tasks and accelerate inference speed. As shown in Figure 18, the IMDM recurrently refines the LR image in a "coarse to fine" manner, and it consists of a 1×1 convolutional layer and several stacked information multi-distillation blocks (IMDB). Benefit from the RNN structure, the proposed network is lightweight but efficient. The final model was trained with L1 loss, VGG perceptual loss, GAN loss, total variation loss, and the novel Fourier spectrum loss [56].

KU_ISPL_A

Inspired by SRFBN [26] and GMFN [24], KU_ISPL_A proposed a recurrent transmission network (see Figure 19) which gradually grows the resolution for each step in the recurrent structure and uses the RDB structure of the RDN [58] model to extract features for each resolution. Each step produces $\times2$, $\times4$, $\times8$ and $\times16$ resolution results.

CET_CVLab

CET_CVLab proposed a V-Stacked relativistic GAN for extreme SR. In the generator part of the network, there are 3 stacks of deep CNN based structure which is inspired from

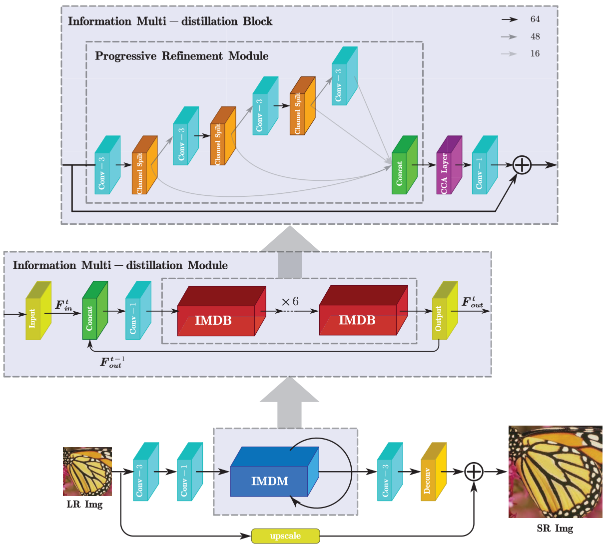


Figure 18. sysu-AIR's network architecture.

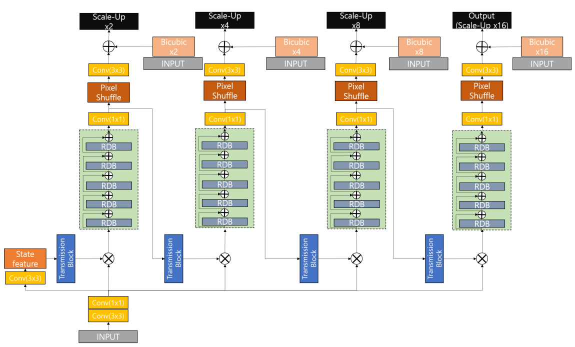


Figure 19. KU_ISPL_A's network architecture.

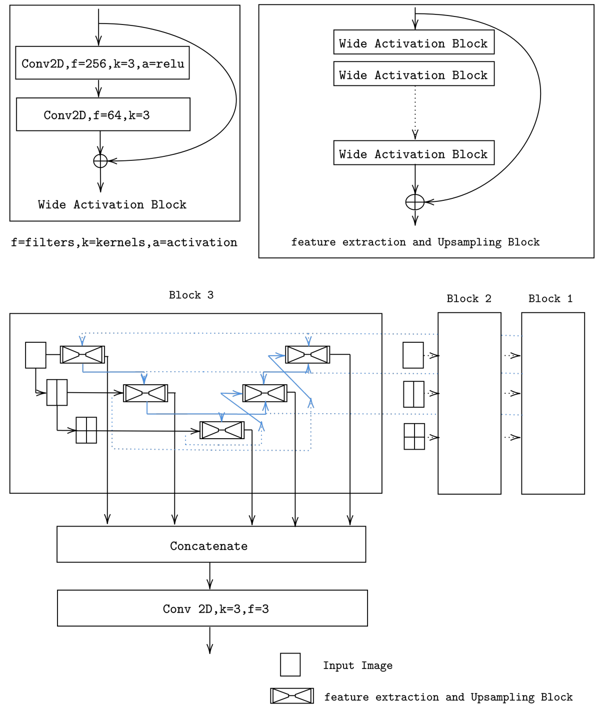


Figure 20. CET_CVLab's network architecture.

[50]. Within each stack there is a pyramidical arrangement of layers that forms the V-shape as shown in Figure 20. Each level of the structure is an encoder-decoder block. There are 5 levels in the pyramid and the initial level takes the whole image as a single patch followed by layers with patches up to 4 and then back to a single patch. All these image patches are passed through a feature extraction network followed by $\times16$ upsampling layer. The generator network is first trained with L1 loss, and then refined with relativistic average GAN loss and VGG perceptual loss.

Acknowledgements

We thank the NTIRE 2020 sponsors: HUAWEI, OPPO, Voyage81, MediaTek, DisneyResearch|Studios, and Computer Vision Lab (CVL) ETH Zurich.

A. Teams and affiliations

NTIRE2020 team

Title: NTIRE 2020 Perceptual Extreme Super-Resolution Challenge

Members:
Kai Zhang (kai.zhang@vision.ee.ethz.ch),
Shuhang Gu (shuhang.gu@vision.ee.ethz.ch),
Radu Timofte (radu.timofte@vision.ee.ethz.ch)

Affiliations:
Computer Vision Lab, ETH Zurich, Switzerland

OPPO-Research

Title: Perceptual Extreme Super-Resolution Network with Receptive Field Block

Members: *Taizhang Shang (shangtaizhang@oppo.com)*, Qiuju Dai, Shengchen Zhu, Tong Yang, Yandong Guo

Affiliation:
OPPO Research

CIPLAB

Title: Investigating Loss Functions for Extreme Super-Resolution

Members: *Younghyun Jo*[1] *(yh.jo@yonsei.ac.kr)*, Sejong Yang[1], Seon Joo Kim[1,2]

Affiliation:
[1] Yonsei University
[2] Facebook

HiImageTeam

Title: Cascade SR-GAN for Extreme Super-Resolution
Members: *Lin Zha*[1]
(zhalin@hisense.com), Jiande Jiang[1], Xinbo Gao[2], Wen Lu[2]
Affiliation:
[1] Qingdao Hi-image Technologies Co.,Ltd (Hisense Visual Technology Co.,Ltd.)
[2] Xidian University

ECNU

Title: Two-Stages-SRGAN
Members: *Jing Liu*
(51174500035@stu.ecnu.edu.cn)
Affiliation:
Multimedia and Computer Vision Lab, East China Normal University (ECNU)

SIA

Title: Perception-Oriented Extreme Upscaling using ESRGAN
Members: *Kwangjin Yoon*
(yoon28@si-analytics.ai), Taegyun Jeon
Affiliation:
SI Analytics Co., Ltd., 441 Expo-ro, Yuseong-gu, Daejeon, 34051, Republic of Korea

TTI

Title: Deep Back Projection for Perceptual Extreme Super-Resolution
Members: *Kazutoshi Akita*
(sd19401@toyota-ti.ac.jp), Takeru Ooba, Norimichi Ukita
Affiliation:
Toyota Technological Institute (TTI)

DeepBlueAI

Title: Bag of Tricks for Perceptual Extreme Super-Resolution ***Members:*** *Zhipeng Luo*
(luozp@deepblueai.com), Yuehan Yao, Zhenyu Xu
Affiliation:
DeepBlue Technology (Shanghai) Co.,Ltd

APTX4869

Title: Progressive Super-Resolving and Refining
Members: *Dongliang He*[1]
(hedlcc@126.com), Wenhao Wu[2], Yukang Ding[1], Chao Li[1], Fu Li[1], Shilei Wen[1]
Affiliation:
[1] Department of Computer Vision Technology (VIS), Baidu Inc.
[2] Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, China.

CNDP-Lab

Title: Cascaded U-Net with Channel Attention for Image Super-Resolution
Members: *Jianwei Li*[1,2]
(ljwdust@gmail.com)
Affiliation:
[1] Peking University
[2] State Key Laboratory of Digital Publishing Technology, Founder Group

MSMers

Title: Progressively Upsampled Residual Channel Attention Network for Extreme Super-Resolution
Members: Fuzhi Yang[1] (yfzcopy0702@sjtu.edu.cn), Huan Yang[2](huayan@microsoft.com), Jianlong Fu[2]
Affiliation:
[1] Shanghai Jiao Tong University
[2] Microsoft Research, Beijing, P.R. China

kaws

Title: Wavelet Pyramid Generation based High-frequency Recovery for Perceptual Extreme Super-Resolution
Members: *Byung-Hoon Kim*[1]
(egyptdj@kaist.ac.kr), JaeHyun Baek[2], Jong Chul Ye[1]
Affiliation:
[1] Korea Advanced Institute of Science and Technology (KAIST)
[2] Amazon Web Services

UIUC-IFP

Title: Progressive WDSR
Members: *Yuchen Fan*
(yuchenf4@illinois.edu), Thomas S. Huang
Affiliation:
University of Illinois at Urbana-Champaign

KU_ISPLB

Title: FBRNN:feedback recurrent neural network
Members: *Junyeop Lee*
(jylee@ispl.korea.ac.kr), Bokyeung Lee, Jungki Min, Gwantae Kim, Kanghyu Lee, Jaihyun Park
Affiliation:
Korea University

MsSrModel

Title: Multi-scale SR Model
Members: *Mykola Mykhailych*

(*mykolam@wix.com*)
Affiliation:
Wix.com Ltd.

MoonCloud

Title: Multi-scale ResNet
Members: *Haoyu Zhong*[1] (*hy0421@outlook.com*), Yukai Shi[1], Xiaojun Yang[1], Zhijing Yang[1], Liang Lin[2]
Affiliation:
[1] Guangdong University of Technology
[2] Sun Yat-sen University

SuperT

Title: Fast and Balanced Laplacian Pyramid Networks for Progressive Image Super-Resolution
Members: *Tongtong Zhao* (*daitoutiere@gmail.com*), Jinjia Peng, Huibing Wang
Affiliation:
Dalian Maritime University

sysu-AIR

Title: A Fast Feedback Network for Large Scale Image Super-Resolution
Members: *Zhi Jin* (*jinzh26@mail.sysu.edu.cn*), Jiahao Wu, Yifu Chen, Chenming Shang, Huanrong Zhang
Affiliation:
School of Intelligent Systems Engineering, Sun Yat-sen University.

KU_ISPL_A

Title: Recurrent Transmission Network for Extreme Super Resolution
Members: *Jeongki Min* (*jkmin@ispl.korea.ac.kr*), Junyeop Lee, Bokyeung Lee, Jaihyun Park, Gwantae Kim, Kanghyu Lee
Affiliation: Korea University

CET_CVLab

Title: Perceptual Extreme Super resolution Using V-Stacked Relativistic GAN
Members: *Hrishikesh P S* (*hrishikeshps@cet.ac.in*), Densen Puthussery, Jiji C V
Affiliation:
College of Engineering Trivandrum

References

[1] Abdelrahman Abdelhamed, Mahmoud Afifi, Radu Timofte, Michael 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*, June 2020. 2

[2] Eirikur Agustsson and Radu Timofte. Ntire 2017 challenge on single image super-resolution: Dataset and study. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, July 2017. 3

[3] Codruta O. Ancuti, Cosmin Ancuti, Florin-Alexandru Vasluianu, Radu Timofte, et al. Ntire 2020 challenge on non-homogeneous dehazing. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2020. 2

[4] Boaz Arad, Radu Timofte, Yi-Tun Lin, Graham Finlayson, Ohad Ben-Shahar, et al. Ntire 2020 challenge on spectral reconstruction from an rgb image. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2020. 2

[5] Yochai Blau, Roey Mechrez, Radu Timofte, Tomer Michaeli, and Lihi Zelnik-Manor. The 2018 pirm challenge on perceptual image super-resolution. In *The European Conference on Computer Vision (ECCV) Workshops*, 2018. 1, 2, 4

[6] Andrew Brock, Jeff Donahue, and Karen Simonyan. Large scale gan training for high fidelity natural image synthesis. *arXiv preprint arXiv:1809.11096*, 2018. 1

[7] Marcel Christoph Bühler, Andrés Romero, and Radu Timofte. Deepsee: Deep disentangled semantic explorative extreme super-resolution, 2020. 1

[8] Jianrui Cai, Shuhang Gu, Radu Timofte, and Lei Zhang. Ntire 2019 challenge on real image super-resolution: Methods and results. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2019. 1

[9] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Image super-resolution using deep convolutional networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(2):295–307, 2015. 1

[10] Dario Fuoli, Zhiwu Huang, Martin Danelljan, Radu Timofte, et al. Ntire 2020 challenge on video quality mapping: Methods and results. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2020. 2

[11] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680, 2014. 1, 4

[12] Shuhang Gu, Martin Danelljan, Radu Timofte, et al. Aim 2019 challenge on image extreme super-resolution: Methods and results. In *International Conference on Computer Vision Workshop (ICCVW)*, pages 3556–3564. IEEE, 2019. 1, 3, 4

[13] Shuhang Gu, Andreas Lugmayr, Martin Danelljan, Manuel Fritsche, Julien Lamour, and Radu Timofte. Div8k: Diverse 8k resolution image dataset. In *International Conference on Computer Vision Workshop (ICCVW)*, pages 3512–3516. IEEE, 2019. 2, 3

[14] Muhammad Haris, Gregory Shakhnarovich, and Norimichi Ukita. Deep back-projection networks for super-resolution. In *IEEE conference on computer vision and pattern recognition*, pages 1664–1673, 2018. 6, 8

[15] Zheng Hui, Xinbo Gao, Yunchu Yang, and Xiumei Wang. Lightweight image super-resolution with information multi-distillation network. In *Proceedings of the 27th ACM International Conference on Multimedia (ACM MM)*, pages 2024–2032, 2019. 8

[16] Andrey Ignatov, Radu Timofte, Thang Van Vu, Tung Minh Luu, Trung X Pham, Cao Van Nguyen, Yongwoo Kim, Jae-Seok Choi, Munchurl Kim, Jie Huang, et al. Pirm challenge on perceptual image enhancement on smartphones: Report. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 0–0, 2018. 1

[17] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In Francis Bach and David Blei, editors, *International Conference on Machine Learning*, volume 37, pages 448–456, 2015. 6

[18] Younghyun Jo, Sejong Yang, and Seon Joo Kim. Investigating loss functions for extreme super-resolution. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2020. 4

[19] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In *European conference on computer vision*, pages 694–711. Springer, 2016. 2, 4

[20] Alexia Jolicoeur-Martineau. The relativistic discriminator: a key element missing from standard gan. *arXiv preprint arXiv:1807.00734*, 2018. 2

[21] Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 4401–4410, 2019. 1

[22] Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, et al. Photo-realistic single image super-resolution using a generative adversarial network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4681–4690, 2017. 1, 8

[23] Junyeop Lee, Jaihyun Park, Kanghyu Lee, Jeongki Min, Gwantae Kim, Bokyeung Lee, Bonhwa Ku, David K. Han, and Hanseok Ko. FBRNN: Feedback recurrent neural network for extreme image super-resolution. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2020. 7

[24] Qilei Li, Zhen Li, Lu Lu, Gwanggil Jeon, Kai Liu, and Xiaomin Yang. Gated multiple feedback network for image super-resolution. *preprint arXiv:1907.04253*, 2019. 8

[25] Yawei Li, Shuhang Gu, Christoph Mayer, Luc Van Gool, and Radu Timofte. Group sparsity: The hinge between filter pruning and decomposition for network compression. In *Proceedings of the IEEE International Conference on Computer Vision*, 2020. 1

[26] Zhen Li, Jinglei Yang, Zheng Liu, Xiaomin Yang, Gwanggil Jeon, and Wei Wu. Feedback network for image super-resolution. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3867–3876, 2019. 8

[27] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. Enhanced deep residual networks for single image super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 136–144, 2017. 7

[28] Songtao Liu, Di Huang, et al. Receptive field block net for accurate and fast object detection. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 385–400, 2018. 4

[29] Ilya Loshchilov and Frank Hutter. SGDR: stochastic gradient descent with restarts. *CoRR*, abs/1608.03983, 2016. 6

[30] Andreas Lugmayr, Martin Danelljan, Radu Timofte, et al. Ntire 2020 challenge on real-world image super-resolution: Methods and results. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2020. 2

[31] Andreas Lugmayr, Martin Danelljan, Radu Timofte, Manuel Fritsche, Shuhang Gu, et al. Aim 2019 challenge on real-world image super-resolution: Methods and results. In *International Conference on Computer Vision Workshop (ICCVW)*. IEEE, 2019. 1

[32] Chao Ma, Chih-Yuan Yang, Xiaokang Yang, and Ming-Hsuan Yang. Learning a no-reference quality metric for single-image super-resolution. *Computer Vision and Image Understanding*, 158:1–16, 2017. 2

[33] Seungjun Nah, Sanghyun Son, Radu Timofte, Kyoung Mu Lee, et al. Ntire 2020 challenge on image and video deblurring. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2020. 2

[34] Youzin Qiu, Ruxin Wang, Dapeng Tao, and Jun Cheng. Embedded block residual network: A recursive restoration model for single-image super-resolution. *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 4179–4188, 2019. 6

[35] 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. 2, 7

[36] Edgar Schönfeld, Bernt Schiele, and Anna Khoreva. A u-net based discriminator for generative adversarial networks. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2020. 2, 5, 6

[37] Taizhang Shang, Qiuju Dai, Shengchen Zhu, Tong Yang, and Yandong Guo. Perceptual extreme super-resolution network with receptive field block. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2020. 4

[38] 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 *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1874–1883, 2016. 4, 8

[39] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014. 2

[40] Radu Timofte, Eirikur Agustsson, Luc Van Gool, Ming-Hsuan Yang, Lei Zhang, et al. Ntire 2017 challenge on single image super-resolution: Methods and results. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, July 2017. 1, 3

[41] Radu Timofte, Vincent De Smet, and Luc Van Gool. A+: Adjusted anchored neighborhood regression for fast super-resolution. In *Asian Conference on Computer Vision*, 2014. 1

[42] Radu Timofte, Shuhang Gu, Jiqing Wu, and Luc Van Gool.

Ntire 2018 challenge on single image super-resolution: Methods and results. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2018. 1

[43] Radu Timofte, Rasmus Rothe, and Luc Van Gool. Seven ways to improve example-based single image super resolution. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016. 3

[44] Xintao Wang, Ke Yu, Chao Dong, and Chen Change Loy. Recovering realistic texture in image super-resolution by deep spatial feature transform. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 606–615, 2018. 3

[45] Xintao Wang, Ke Yu, Shixiang Wu, Jinjin Gu, Yihao Liu, Chao Dong, Yu Qiao, and Chen Change Loy. Esrgan: Enhanced super-resolution generative adversarial networks. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 0–0, 2018. 1, 2, 4, 6

[46] Zhou Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, April 2004. 2, 4

[47] T. Xie, X. Yang, Y. Jia, C. Zhu, and X. LI. Adaptive densely connected single image super-resolution. In *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*, pages 3432–3440, 2019. 5

[48] Jiahui Yu, Yuchen Fan, Jianchao Yang, Ning Xu, Xinchao Wang, and Thomas S Huang. Wide activation for efficient and accurate image super-resolution. *arXiv preprint arXiv:1808.08718*, 2018. 6, 7

[49] 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

[50] Hongguang Zhang, Yuchao Dai, Hongdong Li, and Piotr Koniusz. Deep stacked hierarchical multi-patch network for image deblurring. *CoRR*, abs/1904.03468, 2019. 9

[51] Kai Zhang, Shuhang Gu, Radu Timofte, et al. Aim 2019 challenge on constrained super-resolution: Methods and results. In *IEEE International Conference on Computer Vision Workshops*, 2019. 1

[52] Kai Zhang, Shuhang Gu, Radu Timofte, et al. Ntire 2020 challenge on perceptual extreme super-resolution: Methods and results. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2020. 2

[53] Kai Zhang, Luc Van Gool, and Radu Timofte. Deep unfolding network for image super-resolution. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2020. 1

[54] Kai Zhang, Wangmeng Zuo, and Lei Zhang. Deep plug-and-play super-resolution for arbitrary blur kernels. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1671–1681, 2019. 1

[55] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 586–595, 2018. 2, 4, 5

[56] Xu Zhang, Svebor Karaman, and Shih-Fu Chang. Detecting and simulating artifacts in gan fake images. In *2019 IEEE International Workshop on Information Forensics and Security (WIFS)*, pages 1–6, Dec 2019. 8

[57] Yulun Zhang, Kunpeng Li, Kai Li, Lichen Wang, Bineng Zhong, and Yun Fu. Image super-resolution using very deep residual channel attention networks. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 286–301, 2018. 1, 6, 7

[58] 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. 8