NTIRE 2019 Challenge on Real Image Super-Resolution: Methods and Results

Jianrui Cai Shuhang Gu Radu Timofte Lei Zhang Xiao Liu Yukang Ding
Dongliang He Chao Li Yi Fu Shilei Wen Ruicheng Feng Jinjin Gu
Yu Qiao Chao Dong Dongwon Park Se Young Chun Sanghoon Yoon
Junhyung Kwak Donghee Son Syed Waqas Zamir Aditya Arora Salman Khan
Fahad Shahbaz Khan Ling Shao Zhengping Wei Lei Liu Hong Cai Darui Li
Fujie Gao Zheng Hui Xiumei Wang Xinbo Gao Guoan Cheng Ai Matsune
Qiuyu Li Leilei Zhu Huaijuan Zang Shu Zhan Yajun Qiu Ruxin wang
Jiawei Li Yongcheng Jing Mingli Song Pengju Liu Kai Zhang Jingdong Liu
Jiye Liu Hongzhi Zhang Wangmeng Zuo Wenyi Tang Jing Liu
Youngjung Kim Changyeop Shin Minbeom Kim Sungho Kim
Pablo Navarrete Michelini Hanwen Liu Dan Zhu Xuan Xu Xin Li Furui Bai
Xiaopeng Sun Lin Zha Yuanfei Huang Wen Lu Yanpeng Cao Du Chen
Zewei He Sun Anshun Siliang Tang Hongfei Fan Xiang Li Guo Li
Wenjie Zhang Yumei Zhang Qingwen He Jinghui Qin Lishan Huang Yukai Shi
Pengxu Wei Wushao Wen Liang Lin Jun Yu Guochen Xie Mengyan Li
Rong Chen Xiaotong Luo Chen Hong Yanyun Qu Cuihua Li Zhi-Song Liu
Li-Wen Wang Chu-Tak Li Can Zhao Bowen Li Chung-Chi Tsai
Shang-Chih Chuang Joon-Hee Choi Joonsoo Kim Xiaoyun Jiang Ze Pan
Qunbo Lv Zheng Tan Peidong He

Abstract

This paper reviewed the 3rd NTIRE challenge on single-image super-resolution (restoration of rich details in a low-resolution image) with a focus on proposed solutions and results. The challenge had 1 track, which was aimed at the real-world single image super-resolution problem with an unknown scaling factor. Participants were mapping low-resolution images captured by a DSLR camera with a shorter focal length to their high-resolution images captured at a longer focal length. With this challenge, we introduced a novel real-world super-resolution dataset (RealSR). The track had 403 registered participants, and 36 teams competed in the final testing phase. They gauge the state-of-the-art in real-world single image super-resolution.

1. Introduction

Single image super-resolution (SISR) [13] aims to restore a high-resolution (HR) image from its low-resolution (LR) observation. SISR has been an active research topic for decades [29, 44, 34, 36, 1] because of its high practical values in enhancing image details and textures. Since SISR is a severely ill-posed inverse problem, for each LR image the space of plausible corresponding HR images is huge and scales up quadratically with the magnification factor, learning image prior information from the HR and/or LR exemplar images [13, 10, 42, 15, 11, 3, 20, 43, 8, 16, 35, 31] plays an indispensable role in recovering the details from an LR input image.

Thanks to the rapid development of deep convolutional neural networks (CNNs) [23], recent years have witnessed an explosive spread of training CNN models to perform SISR, and the performance has been consistently improved by designing new CNN architectures [7, 39, 32, 19, 33, 25, 46, 45] and loss functions [17, 24, 30]. Though significant advances have been made, most of the existing SISR methods are trained and evaluated on simulated datasets which

J. Cai (csjcai@comp.polyu.edu.hk, The Hong Kong Polytechnic University), S. Gu, R. Timofte and L. Zhang are the NTIRE 2019 challenge organizers, while the other authors participated in the challenge.

Appendix A contains the authors' teams and affiliations.
NTIRE webpage: http://www.vision.ee.ethz.ch/ntire19/

assume simple and uniform degradation (*i.e.*, bicubic degradation). Unfortunately, SISR models trained on such simulated datasets are hard to generalize to practical applications since the authentic degradations in real-world LR images are much more complex [41, 21].

The NTIRE 2019 challenge takes a step forward in benchmarking example-based real-world single image super-resolution. It uses a 100 real-world low and high resolution image pairs (RealSR dataset [2]) obtained in diverse indoor and outdoor environments. Each image in the RealSR dataset has a pixel resolution no smaller than 1000×1000. These images were captured with the high-end DSLR cameras and lens.

2. NTIRE 2019 Challenge

The objectives of the NTIRE 2019 challenge on example-based real-world single-image super-resolution are: (i) to gauge and push the state-of-the-art in real-world single image super-resolution (SISR); (ii) to compare different solutions; and (iii) to promote a novel large real-world single image super-resolution dataset (RealSR [2]).

2.1. RealSR Dataset

The RealSR dataset consists of HR-LR image pairs captured from the same scene by adjusting the lens of DSLR cameras. Sophisticated image registration operations are also performed to generate the HR and LR pairs of the same content. The detailed dataset construction process can be found in [2].

2.2. Track and competition

Track: Real-world SR. High-resolution images captured by DSLR cameras with a longer focal length and their corresponding low-resolution image captured at a shorter focal length are provided for training. The optical degradation process is the same (use the same way to capture the HR-LR image pairs) within each image space and for all the images in train, validation, and test sets.

Competition. In this track there is a competition for an unknown scaling factor and degradation operators. CodaLab platform was used for this competition of NTIRE 2019 challenge. To access the data and submit their HR image results to the CodaLab evaluation server each participant had to register.

Challenge phases. (1) *Development (training) phase:* the participants got both LR and HR train images and the LR images of the RealSR dataset; (2) *Validation phase:* the participants had the opportunity to test their solutions on the LR validation images and to receive immediate feedback by uploading their results to the server. A validation leaderboard is available; (3) *Final evaluation (test) phase:* the participants got the LR test images and had to submit both their super-resolved image and a description of their methods before the challenge deadline. One week later the final results were made available to the participants.

(a) LR (b) HR (c) SuperRior (d) SuperSR

Figure 1. The super-resolved results by participants.

Evaluation protocol. The Peak Signal-to-Noise Ratio (PSNR) measured in deciBels (dB) and the Structural Similarity index (SSIM) [38] computed between an image result and the ground truth are the quantitative measures. The higher the score is the better the restoration fidelity to the ground truth image. A rim of 6 + s image pixels, where s is the magnification factor, are ignored in the evaluation.

3. Challenge Results

From 403 registered participants, 36 teams entered in the final phase and submitted results, codes/executables, and factsheets. Table 1 reports the final test results, rankings of the challenge, self-reported runtimes and major details from the factsheets. The methods are briefly described in section 4 and the team members are listed in Appendix A.

Architectures and main ideas. All the proposed methods are deep learning based. The very deep residual channel attention networks (RCAN) architecture [45] is the basis for most of the proposed methods. To improve the efficiency, most of the team use multi-scale architecture to downsample and upsample the image representations.

Restoration fidelity. The top 4 methods (SuperRior, SuperSR, BMIPL_UNIST_DW and David_MM_AI) achieved similar PSNR scores (within 0.07dB). IVIP-LAB entry, ranked 9th, is only 0.21dB behind the best PSNR score of SuperRior. The super-resolved results by SuperRior and SuperSR can be found in Figure 1.

Ensembles and fusion. Data augmentation is commonly used technique for all these participants. Most teams employ pseudo-ensembles [37]. The inputs are flipped/rotated and the HR results are aligned and averaged for enhanced prediction.

Conclusions. By analyzing the settings, the proposed methods and their results we conclude: (i) The proposed methods improve the state-of-the-art in real-world SISR. (ii) SISR models trained on simulated data (*i.e.*, bicubic downsampling) are hard to generalize to practical applications since the authentic degradations in real-world LR images are much more complex. (iii) The proposed RealSR dataset

Table 1. NTIRE 2019 Real-world SR Challenge results, final rankings, runtimes [s] per test image and details from the factsheets.

Team	Author	PSNR	SSIM	runtime [s]	Platform	CPU/GPU (at runtime)	Ensemble
SuperRior	dingyukang	29.00	0.84	60.00	Pytorch	Tesla V100	flip/rotation (×8)
SuperSR	jnjaby	28.97	0.84	36.90	Pytorch	GTX 1080	flip/rotation (×8)
BMIPL_UNIST_DW	Sprite	28.93	0.84	600.00	Pytorch	Titan V	flip/rotation (×8)
David_MM_AI	sean.shy	28.93	0.83	106.56	TensorFlow	Tesla V100	flip/rotation (×8)
LulluVision	CountVladimir	28.88	0.84	38.00	Pytorch	GTX 1080ti	flip/rotation (×8)
TeamInception	swz30	28.88	0.83	98.00	Pytorch	Titan Xp	flip/rotation (×8)
LiveMe_AILab	hanshui	28.87	0.83	–	Pytorch	Quadro P6000	flip/rotation (×8)
rainbow	zheng222	28.81	0.83	10.00	Pytorch	GTX 1080Ti	flip/rotation (×8)
IVIP-LAB	GCheng	28.79	0.84	47.08	Pytorch	GTX 1080Ti	flip/rotation (×8)
Future	qiuyj	28.76	0.83	100.28	Pytorch	Titan X	flip/rotation (×8)
YZSR	ycjing	28.73	0.83	3.00	TensorFlow	Tesla P100	flip/rotation (×10)
HIT-UltraVision	UltraVision	28.72	0.83	12.54	Pytorch	GTX 1080Ti	flip/rotation (×8)
Meteor	loseall	28.63	0.83	225.00	Tensorflow/PyTorch	Titan X	flip/rotation (×8)
ECNU	exciting	28.61	0.83	4.09	Pytorch	GTX 1080Ti	flip/rotation (×8)
AiDDle	CGOTY	28.60	0.83	360.00	Pytorch	Titan Xp	flip/rotation (×8)
BOE-IOT-AIBD	BOE-IOT-AIBD	28.54	0.83	1122.00	Pytorch	Tesla P100	-
xuxu123	xuxu123	28.53	0.83	350.00	Pytorch	Titan Xp	flip/rotation (×8)
VIPSL	vipsl	28.52	0.83	–	Pytorch	GTX 1080ti	flip/rotation (×8)
AP_FStone	LONGTAOPENG	28.48	0.83	60.00	TensorFlow	Tesla P40	-
ZXHresearch [12]	shangqigao	28.46	0.83	21.06	-	-	-
zju231	zeweihe	28.45	0.83	4.35	MatConvNet	Quadro P6000	flip/rotation (×4)
KSC	fanhongfei	28.44	0.82	3.00	TensorFlow	-	-
Shuan	Shuan	28.41	0.82	–	-	-	-
MENet	scape1989	28.35	0.83	380.31	Pytorch	GTX 1080ti	flip/rotation (×8)
pws_SR	wangdaru	28.24	0.83	1.64	TensorFlow	Tesla P40	-
Early	Early	28.21	0.82	60.00	Pytorch	GTX 2080	-
egg126	egg126	28.19	0.82	4.45	Pytorch	GTX 1080ti	flip/rotation (×4)
XMU_IMAGE	sudo	28.18	0.82	26.14	Pytorch	GTX 1080	N/A
DeepSR	ZSLiu	28.15	0.82	30.00	Caffe	GTX 1080Ti	flip/rotation (×8)
Pikachu&Sparky&Co.	lbwdruid	28.06	0.82	200.00	TensorFlow	Tesla K40	flip/rotation (×8)
QCAM	chungchi	27.89	0.81	2.00	TensorFlow	Tesla K80	N/A
LAN	emma2019chen	27.78	0.81	6.58	-	-	-
ironhead	ironhead	27.26	0.80	2.10	-	-	-
NEWBEE	pz102	27.10	0.79	4.53	TensorFlow	GTX 1080Ti	-
jinxin	chenzhikai	27.03	0.78	0.21	-	-	-
late_autumn	huangweijian	24.70	0.78	–	-	-	-
Baseline	Bicubic	26.89	0.78	–	-	-	-

is highly desired for learning-based methods to deal with the real-world SISR problem.

4. Challenge Methods and Teams

4.1. SuperRior team utilized a u-net architecture for SISR. They thus call the model U-shaped Deep Super-Resolution (UDSR), and illustrated their architecture in Figure 2. In UDSR, the input low-resolution image is operated by a convolution layer to extract deep feature maps. They continued to process the feature maps by residual blocks and downsample them to lower resolution. They then upsampled the feature maps and apply additional residual blocks to obtain high-resolution feature maps. They built straightforward paths from the left part of the U-shaped network to the right part of the network. They used the feature maps of the highest resolution to output a residual image and add it to the input image as the final output.

Besides, they train models in an unified three-stage cascaded framework such that the input image could be refined better in each stage (Figure 3). In each stage, they use the output of the previous stage as input and use an UDSR to process the input image. For three stages have different supervision signals, from coarse to fine. For the first stage, they downsample the high-resolution ground-truth by 4 scales, and then upsample it to the original size. The 4× blurred image is utilized to supervise the output of the first UDSR model. In the second stage, the 2× blurred image is utilized to supervise the output of the second UDSR model. In the third stage, the ground-truth image is used to supervised the output of the third UDSR model.

Finally, they utilized an adaptive multi-model ensemble method for results fusion. Model diversity exists among different models. They notice that even with the same model, performances of different patches vary a lot. These priors motivate them to ensemble multiple models in an adaptive way, namely, the fusion weights of different models must be conditioned on the frames generated by these models in

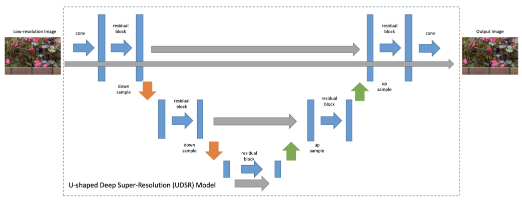

Figure 2. The overall architecture of SuperRior team.

Figure 3. The cascaded learning framework of SuperRior team.

a image patch granularity. They use a CNN model to operate the outputs of several different models, and learn a normalized weight for each pixel of each single model.

4.2. SuperSR team investigated underlying solutions on real single-image super-resolution task and proposed a modified U-Net architecture with several regularization techniques applied to combat the potential for over-fitting [9]. Despite the use of data augmentation (*e.g.*, random crop, rotation, flipping), large models exhibit undesirable over-fitting behaviors due to the limited data, leading to poor generalization on the validation set. For instance, a well-trained model tends to generate unpleasant artifacts on restored images. To alleviate this problem, they propose a simple yet effective method to collect additional training data. They first model the image degradation process using a network of the same architecture as described before, with the main difference being that this network takes the HR image as input and generate the LR counterpart. These two models are denoted as D network (Degradation) and R network (Regression), respectively. The D network is trained with provided 60 training image pairs, after which they obtain synthetic LR images on DIV2K training set through a well-optimized D network. To train R network, they treat the synthetic data as additional training data and mix them into the original real data and do not distinguish between the two during training. They also utilize the mixup method to further bridge the gap between real data and synthetic data. The overall architecture can be found in Figure 4.

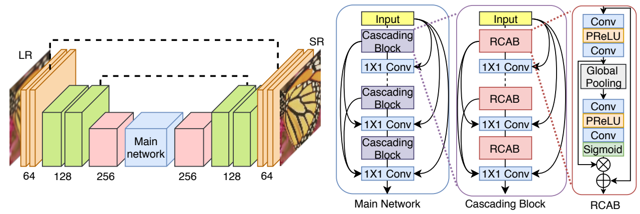

Figure 4. The overall architecture of SuperSR team.

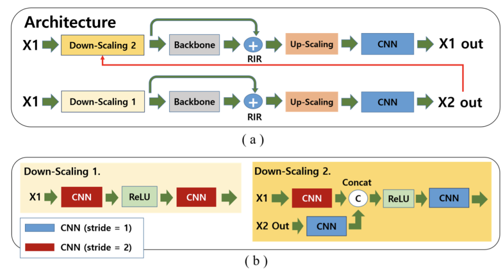

Figure 5. The overall architecture of BMIPL_UNIST_DW team.

4.3. BMIPL_UNIST_DW team created a network with a module base [28]. Figure 5 is shown their network architecture. Their network firstly down-scales the input color image with the size of W $\times$ H $\times$ 3 to the feature maps with the size of W/4 $\times$ H/4 $\times$ 64 using the "Down-Scaling 1" module in Figure 5 (b). Then, these feature maps are fed into the backbone network as well as residual in residual (RIR) skip connection to yield initial Real Super-Resolution feature maps in the spatial dimension of W/4 $\times$ H/4. Then, upscaling and CNN will result in the intermediate Real Super-Resolution image estimate with the size of W/2 $\times$ H/2 $\times$ 3. This result is combined with the down-scaled image in the "Down-Scaling 2" module in Figure 5 (b) for the deblurring at the scale of W/2 $\times$ H/2 using another backbone network, RIR, and up-scaling process to yield the final deblurred output image with the size of W $\times$ H. They used residual channel attention network (RCAN) [45] as a backbone of their network.

4.4. David_MM_AI team strove to build more efficient basic building block. The basic building block of their network is Dual Path Channel Attention Block (DPCAB), which is motivated by DPN [5] and RCAN [45]. Each DPCAB contains 5-layer dual path block (DPB), which contains residual alike path and densely connected alike path. DPB is followed by channel attention (CA) block. As shown in Figure 6, they stacked DPCAB in a structure of residual in residual (RIR), which is motivated by RCAN. Residual Group (RG) contains multiple DPCAB followed by convolutional layer. Their final network, dual path channel attention network (DPCAN), is constructed by stacking

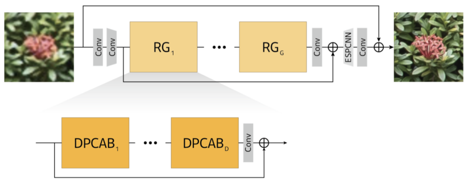

Figure 6. The overall architecture of David_MM_AI team.

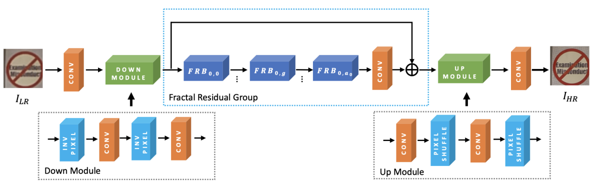

Figure 7. The overall architecture of LulluVision team.

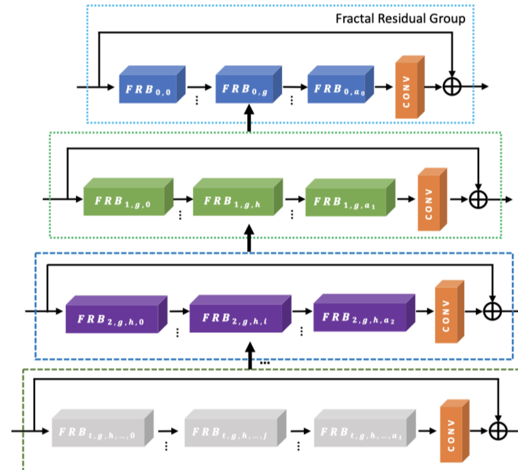

Figure 8. The architecture of fractal residual group by LulluVision team.

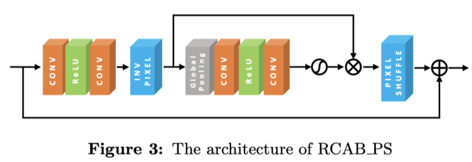

Figure 9. The architecture of RCAB_PS by LulluVision team.

multiple RGs.

4.5. LulluVision team proposed fractal residual network which takes interpolated LR image as input [22]. The overall architecture of fractal residual network is shown in Figure 7. They introduced the down-up module for efficient memory usage. The down-up module is a combination of the downsampling module at the front of the model and the upsampling module at the back of the model. The downsampling module reduces the size of shallow feature map to lessen usage of memory. The downsampling module is composed of two convolution layers and two inverse pixel shuffling layers. The feature map is upsampled at the end of model, using upsampling module. The upsampling module consists of two convolution operation and two pixel shuffling layers. In this way, they can make deeper network under the same GPU memory condition because the downsampling module reduces the size of feature map.

Besides, they proposed the auto-encoder loss, optimized with ℓ_1 loss function generally used in many SR models. The auto-encoder loss makes the down-up module work well. And they also proposed the fractal residual network which has self-similarity like fractal structure. The same network structure is repeated in fractal residual group with different scale as shown in Figure 8. They used the fractal residual network, to make very deep network trainable and improve the performance. They used residual channel attention block with pixel shuffling (RCAB_PS) as basic block of our proposed fractal residual network. RCAB_PS is the extension of residual channel attention to speed up the inference time than RCAB [45]. The architecture of RCAB_PS is shown in Figure 9.

4.6. TeamInception team proposed a deep residual network with spatial and depth-wise attention. The complete framework is shown in Figure 10. Inspired from the work of [45] on super resolution, they proposed a method that is recursive in nature. The main idea is to gradually recover high-frequency components such as edges and textures from the LR input image. This fine-grained detail is added back to the input image (within the network) in order to obtain super-resolved image. Next they describe each component of their recursive dual attention network (RDAN). At the entry point of the RDAN they employ a convolutional layer that takes as input a low resolution image and extracts low-level features. The feature maps become input to the detail decomposition module (DDM), which contains N number of recursive residual groups (RRGs). The goal of RRG is to progressively recover the information related to the desired super-resolved image from the input low-resolution image. Each RRG further employs M number of dual attention blocks (DABs). Features that are less important get suppressed in the DAB, and only useful information is propagated onward. To discern such features, they apply two types of attention mechanisms in DAB: depth-wise attention, and spatial attention.

4.7. LiveMe_AILab team proposed a very deep networks to handle the nonlinear relationship of the LR/HR images.

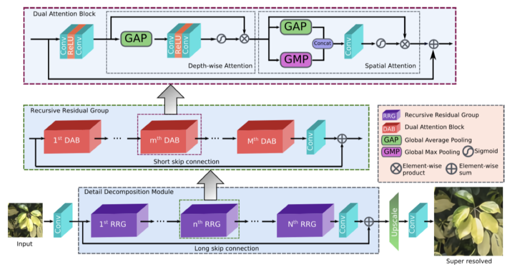

Figure 10. The architecture of TeamInception team.

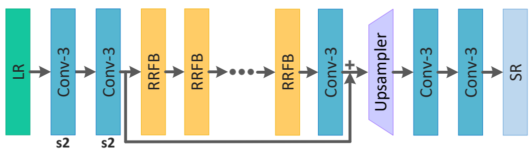

Figure 11. Architecture of our residual fusion network. Here, RRFB denotes residual in residual fusion block, where fusion block is depicted in Figure 12.

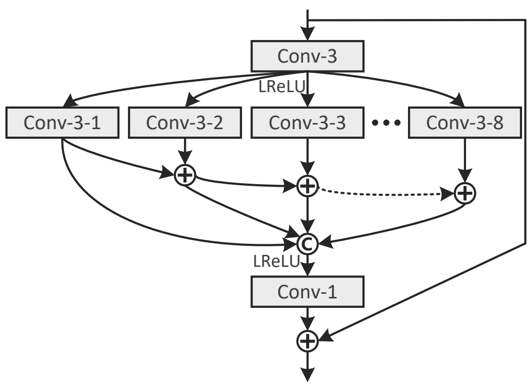

Figure 12. Architecture of proposed fusion block. “Conv-3-3” represents 3×3 convolution layer with dilation rate of 3.

They align the RGB mean of the LR and HR.

4.8. rainbow team proposed a solution based on EDSR structure employs hourglass-like architecture (see Figure 11) to directly predict the HR image. They use the proposed effective fusion block to construct their basic block by residual-in-residual style, which is denoted as residual in residual fusion block (RRFB). Specifically, they use 24 RRFBs for real image super-resolution task. Concerning the input images with arbitrary sizes, they first crop the input to obtain 4 overlapped image patches (their heights and widths can be divisible by 4) and then perform the feed-forward program. The final step is aggregating these patches and producing the entire image corresponding to the input one.

The training procedure contains two stages. They train the network with all training images (“cam1_*.png” and “cam2_*.png”) at the first stage. In the second stage, they train special models for cam1 and cam2, respectively.

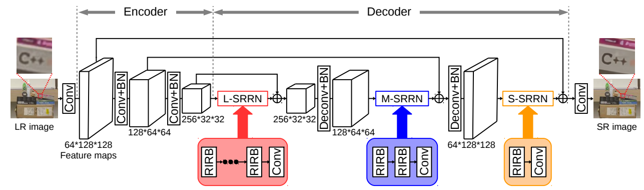

Figure 13. The architecture of IVIP-LAB team.

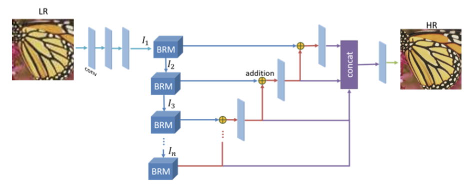

Figure 14. The architecture of Future team.

4.9. IVIP-LAB team proposed an encoder-decoder residual network (EDRN) [6]. They utilized an encoder-decoder structure embedded coarse-to-fine methods. Their encoder-decoder structure upscales/downscales twice for capturing larger receptive field to extract features with more context information. They embedded the coarse-to-fine structure to restore lost information and attenuate the effects of noise gradually. They also discussed the usage of normalization, and further demonstrated that applying batch normalization for downscaling/upscaling convolution layers can reduce the effect of noise and relieved overfitting. Their overall architecture is illustrated in Figure 13: the decoder of their EDRN consists of large-scale residual restoration network (L-SRRN), middle-scale residual restoration network (M-SRRN), and small-scale residual restoration network (S-SRRN). Among them, L-SRRN has four residual in residual blocks (RIRB), M-SRRN has two, and S-SRRN has one. Each RIRB includes 10 residual channel-wise attention blocks (RCAB), one skip connection, and one single convolution layer. RCAB is inherited [45].

4.10. Future team proposed an embedded block residual network (EBRN), as shown in Figure 14, for single image super-resolution.

4.11. YZSR team proposed an enhanced WDSR algorithm with dense multi-branch module (named DMSR), which can addresses the shortcoming of WDSR. The network architecture of DMSR is depicted in Figure 15.

4.12. HIT-UltraVision team proposed a multi-level wavelet residual network (MWRN) by incorporating MWCNN [27] with multiple residual blocks [18] in each level of the encoder and decoder (see Figure 16). Similar to MWCNN, MWRN adopts discrete wavelet transform

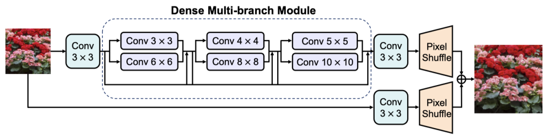

Figure 15. The architecture of YZSR team.

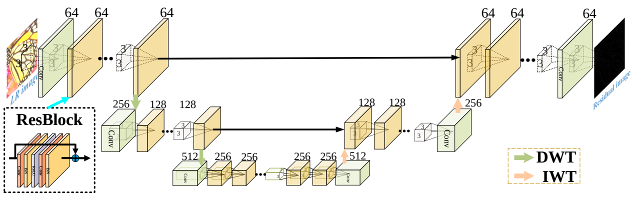

Figure 16. The architecture of HIT-UltraVision team.

(DWT) as the downsamping operator in the second and third level of the encoder and inverse wavelet transform (IWT) as upsampling operator in the corresponding decoder. Different from MWCNN, MWRN uses several residual blocks in each level to enhance feature representation and ease the training. The number of residual blocks in each level is set to 12, while the numbers of feature maps in the three levels are set to 64, 128 and 256, respectively. In order to compress the feature maps after DWT and expand after IWT, additional 3×3 convolution is deployed. In the training phase, the Adam algorithm and ℓ_2-norm loss function are utilized to optimize the network parameters.

4.13. Meteor team proposed a cascaded residual dense network (see Figure 17). The model can be divided into 2 components. In the encoder, cascaded block (CB) following convolution layer with a stride of 2 is responsible for downsampling image features from the previous level to the next one. While in the decoder, cascade block and nearest up-sampler following 2 convolution layers are responsible for recovering the image features. Skip connections are also introduced to avoid losing textural details. There are 3 CBs in both encoder and decoder, and 2 skip connections are concatenated after the nearest up-sampler (as marked in the gray dash box), while 1 long residual skip connection element-wisely adds early extraction features up to the final features. They also constructed the cascade block as a cascaded residual dense block (CRDB), as marked in the orange dash box. They used 3 small residual dense blocks (RDB, as marked in the red dash box) to form a CRDB. 1×1 convolution layer is inserted after each RDB to reduce feature numbers for efficiency. A channel attention layer (CA, marked in the yellow dash box) adapted from RCAN [45] is placed after the last 1×1 convolution to emphasis channel attention for different features.

4.14. ECNU team proposed an end-to-end Encoder-

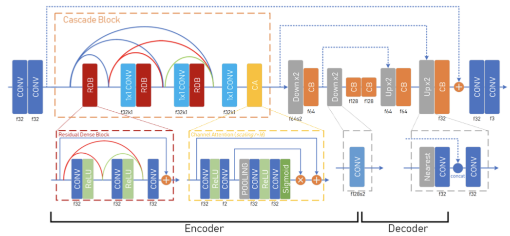

Figure 17. The architecture of Meteor team.

Decoder structure Network(EDN) to learn the SISR mapping function.

4.15. AiDDle team proposed a stacked lossless deconvolutional network (SLDN) using the invertible down-sampling operator (corresponds to the reverse pixel shuffle, *i.e*., spatial resolution is halved while the number of channels increases by a factor of four). In SLDN, a series of shallow deconvolutional networks, which are named as stacked lossless deconvolutional block(SLDB), are stacked to aggregate contextual information and guarantee the recovery of fine details. Single SLDB consists of an encoder module and a corresponding decoder one. In an encoder module, they used two invertible down-sampling layers to enlarge the receptive fields without any information loss, resulting in low-resolution feature maps. A decoder module takes these as input, and produces HR feature representations with the forward pixel shuffle, followed by convolution layers. They further design inner shortcut connections between encoder and decoder to improve gradient flow throughout the network. In that perspective, previous methods treat each channel features equally, which is not flexible enough for the real-world scenario. To make the SLDB focus on more informative features, they introduce an attention mechanism, exploiting the inner-dependencies within feature channels. Specifically, they first describe a channel-wise attention by using global average pooling and gating. The encoder feature maps are then modulated by these attentions, and are directly connected to the subsequent decoder. Finally, the outputs of each SLDB are globally fused to take advantage of the hierarchical representation capability. The SLDN can generate high-quality HR images without noticeable artifacts, as will be confirmed by their results.

4.16. BOE-IOT-AIBD team proposed a new deep learning architecture with the aim to solve general image enhancement problems. Their network design follows signal processing principles based on the Iterative BackProjection (IBP) algorithm (see Figure 18). Compared to similar approaches, they proposed a novel solution to make backprojections run in multiple resolutions by using a data pipeline workflow.

The residual and sequential nature of this system pro-

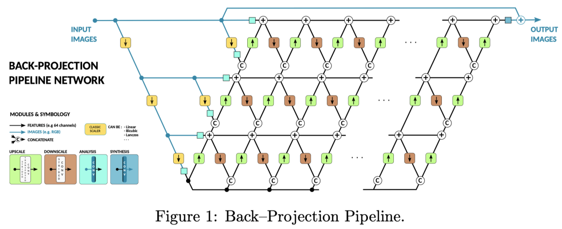

Figure 18. Spatial Color Attention Module presented by BOE-IOT-AIBD team.

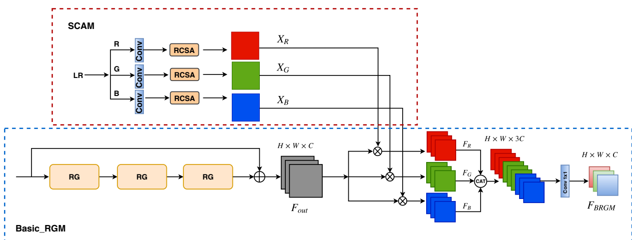

Figure 19. The architecture of xuxu123 team.

vides convenient features similar to ResNets. First, there are straight connection from the output to every convolutional layer, providing gradient superhighways; and second, the sequential processing allows very large models to run efficiently without demanding excessive memory. Finally, a distinctive feature of their design is the use Instance Normalization layers. They found this to be very effective to make the network stable during training. At the expense of better results, the final architecture presents difficulties to be applied when image resolutions are very different in training and inference.

4.17. xuxu123 team developed a novel attention module called Spatial Color Attention Module (SCAM) to jointly exploit the spatial and spectral dependency within color images [40]. Their proposed SCAM module can calibrate important color information for individual color components from output feature maps and help the whole networks to focus on informative features and exploit interdependencies among color channels. SCAM can easily be integrated with other existing image SR networks, as shown in Figure 19.

4.18. VIPSL team introduced the idea of multi-scale channel attention into their model. Compared to the previous channel attention mechanism, their method can learn a richer inter-channel relationship, thereby improving network performance.

4.19. zju231 team proposed the Orientation Attention Network (OAnet) (see in Figure 20) for real image Super-Resolution [4]. The basic building block of OAnet is Orientation Attention Module (OAM), which mainly consists of three directional convolutional operations and a channel attention part [14, 45]. Totally 16 OAMs are employed to learn the LR-to-HR mapping function. To ease the training process, they used extra supervisions in the 4^{th}, 8^{th} and 12^{th} OAM instead of adopting supervision in the final OAM, *i.e.*, the 16^{th} OAM. In order to trade-off the performance and the inference time, pixel shuffle operation [32] was used.

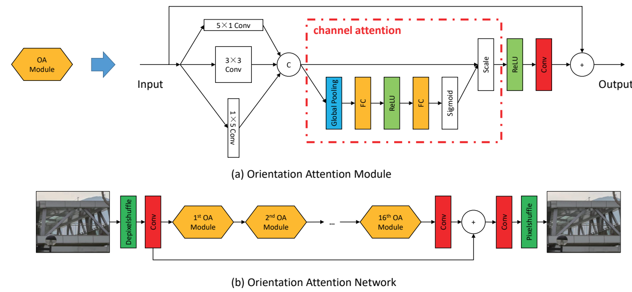

Figure 20. The architecture of zju231 team.

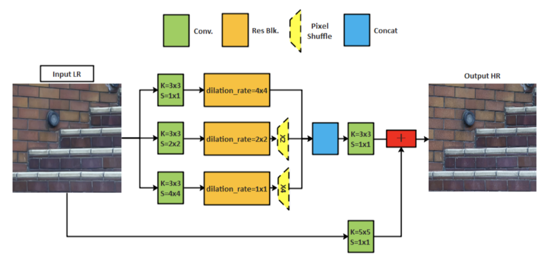

Figure 21. Pyramid based Deep Residual Networks Schema presented by KSC team.

4.20. KSC team modified the WDSR baseline mainly by two aspects. The first part is adding pyramid structure in the WDSR network. And the second part is refining the resblock in the network. They also introduced the SE-block in the resblock and also replaced the normal convolution by dilated convolution. The overall architecture can be found in Figure 21.

4.21. MENet team modified RCAN [45] to adapt this real super-resolution track. The main structure of their solution is shown in Figure 22. Their solution mainly consists three parts: shallow feature extraction, deep feature extraction, reconstruction part. In shallow feature extraction, they use $3{\times}3{\times}64$ convolution with stride 1 to extract shallow feature from the LR input. Then they use 10 residual groups, which is a residual in residual structure and each group have 20 residual blocks, to extract deep features. After they get the deep feature, their solution will conduct a long residual learning which generate a powerful feature representation by summing shallow feature and deep feature. At last,

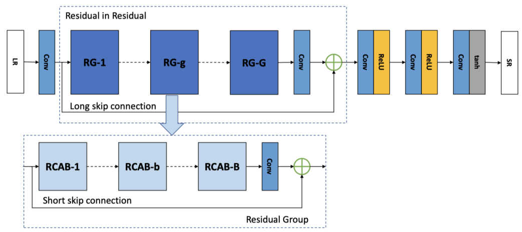

Figure 22. The overall architecture of MENet team.

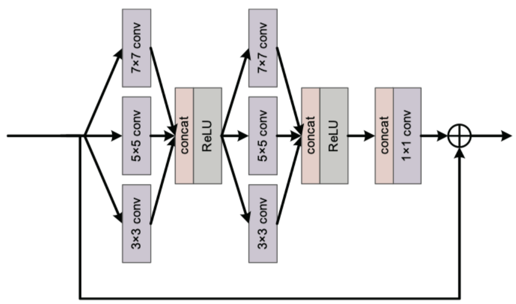

Figure 23. The overall architecture of Early team.

they use three additional layers with kernels of size 3×3 convolution. The first two convolutions are followed by ReLU activation function and the last one is followed by a scaled tanh is applied to the outputs. All inputs are normalized from 0 to 1 by being divided by 255. The whole network can be optimized end-to-end with the standard backpropagation method.

4.22. Early Team proposed a multi-path network (see Figure 23) which utilizes information of different scales to facilitate detail texture synthesis process. Like many other popular methods, Their model also adopt residual structure. However, in each block three paths with different kernel sizes are designed in parallel and concatenated before activation function. 1×1 convolution is also applied to make sure the same size of input and output. The architecture helps to improve low-level feature subtraction efficiency greatly and is beneficial for information transmission.

4.23. egg126 Team proposed a multiple scale residual with spatial and channel attention network for super-resolution.

4.24. XMU_IMAGE Team divided train and validate datasets of competition into two categories, cam1 and cam2. First, they used the SRGAN [24] network to augment the dataset. The input to the generator network is a high-resolution image, and the ground truth is the corresponding low-resolution image. The discriminator network is used to distinguish the generated low-resolution image and the real low-resolution image. Then, modified RCAN [45] with 5 residual groups, 10 residual blocks, 64 feature maps for scaling factor 1× is adopted, and two best models for cam1 and cam2 is trained. Finally, the results of the two models are output separately.

4.25. DeepSR Team proposed a Hierarchical Back Projection Network (HBPN) [26] which adopts the HourGlass (HG) structure to explore deep feature representations for image Super-Resolution (SR). It is made of 3 hourglass SR modules, which called HG-SR blocks, hierarchically reconstruct the fine details of images to generate the final SR image. Each hourglass module contains 6 back projection based blocks to sequentially down- and up-sample the feature maps. In the final stage, they proposed a Weighted Reconstruction (WR) block, which uses the softmax function to normalize the weighting maps, and then combine the weighted outputs of 3 hourglass SR modules to generate the final SR image.

4.26. Pikachu&Sparky&Co. Team proposed effects of Data augmentation on super resolution method.

4.27. QCAM Team proposed a variation of SRCAN [24], which uses the same discriminator and different generator. Their generator has 16 feature maps and 25 residual blocks for each convolution layer while SRGAN has 64 and 16, respectively. Also, the input image of this model is Y channel image (gray scale).

4.28. NEWBEE proposed a multi-scale gate fusion network (MSGFN). The network consists of five modules: deblurring module, half/full size feature extraction modules, gate module and restoration module. After resizing to the half of the original image size, the low resolution image is put into the deblurring module to generate clear image at half size. And it is also the input of the half size feature extraction. Then, the half size features are added with the features from deblurring module, which is multiplied by the output of gate module. Finally, these features are send into the restoration module. Inside the restoration module, the features go through several resblocks, and then are upsampled to full size. After adding with features from full size feature extractor and a few convolutional layers, the restoration image is generated.

Acknowledgements

We thank the NTIRE 2019 sponsors: OPPO Corp., NVIDIA Corp., HUAWEI Technologies Co. Ltd., SAMSUNG Electronics Co. Ltd., Amazon.com, Inc., MEDIATEK Inc., ETH Zurich, and The Hong Kong Polytechnic University.

A. Teams and affiliations

NTIRE2019 team

Title: NTIRE 2019 Challenge on Real Image Super-Resolution

Members: *Jianrui Cai*[1]
(csjcai@comp.polyu.edu.hk), Shuhang Gu[2], Radu Timofte[2], Lei Zhang[1]
Affiliations:
[1]The Hong Kong Polytechnic University
[2]Computer Vision Lab, ETH Zurich, Switzerland

SuperRior

Title: Cascaded U-shaped Deep Super-Resolution
Members:*Xiao Liu*
(liuxiao12@baidu.com), Yukang Ding, Dongliang He, Chao Li, Yi Fu, Shilei Wen
Affiliation:
Computer Vision Technology (VIS) Department, Baidu Inc.

SuperSR

Title: Suppressing Model Overfitting for Image Super-Resolution Networks
Members:*Ruicheng Feng*[1]
(jnjaby@gmail.com), Jinjin Gu[2], Yu Qiao[1,3], Chao Dong[1]
Affiliation:
[1]SIAT-SenseTime Joint Lab, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences
[2]The School of Science and Engineering, The Chinese University of Hong Kong, Shenzhen
[3]The Chinese University of Hong Kong

BMIPL_UNIST_DW

Title: Down-Scaling with Learned Kernels in Multi-Scale Deep Neural Networks for Non-Uniform Single Image Real Super-Resolution
Members:*Dongwon Park*
(dong1@unist.ac.kr), Se Young Chun
Affiliation:
Ulsan National Inst. of Science and Technology(UNIST)

David_MM_AI

Title: Dual Path Channel Attention Networks
Members:*Sanghoon Yoon*
(sean.shy@kakaocorp.com)
Affiliation:
Kakao

LulluVision

Title: Fractal Residual Network and Solutions for Real Super-Resolution
Members:*Junhyung Kwak*[1]
(snu11ee@gmail.com), Donghee Son[2]
Affiliation:
[1]CASTIS Multimedia R&D Center
[2]NALBI Inc.

TeamInception

Title: Image Super Resolution via Recursive Residual Network with Dual Attention
Members:*Syed Waqas Zamir*
(waqas.zamir@inceptioniai.org), Aditya Arora, Salman Khan, Fahad Shahbaz Khan, Ling Shao
Affiliation:
Inception Institute of Artificial Intelligence (IIAI)

LiveMe_AILab

Title: -
Members:*Zhengping Wei*
(weizhengping@cmcm.com), Lei Liu, Hong Cai, Darui Li, Fujie Gao
Affiliation:
LiveMe Inc

rainbow

Title: Residual Fusion Network
Members:*Zheng Hui*
(zheng_hui@aliyun.com), Xiumei Wang, Xinbo Gao
Affiliation:
School of Electronic Engineering Xidian University

IVIP-Lab

Title: Encode-Decode Residual Network
Members:*Guoan Cheng*
(guoan@mail.hfut.edu.cn), Ai Matsune, Qiuyu Li, Leilei Zhu, Huaijuan Zang, and Shu Zhan
Affiliation:
School of Computer Science and Information Engineering, Hefei University of Technology

Future

Title: Embedded Block Residual Network
Members:*Yajun Qiu*
(qyjyun@gmail.com), Ruxin wang
Affiliation:
Yunnan University

YZSR

Title: DMSR: Dense Multi-branch Module for Single Image Super-Resolution

Members:*Jiawei Li*[1]
(mingong.ljw@alibaba-inc.com), Yongcheng Jing[2], Mingli Song[2]
[1]Alibaba Group
[2]College of Computer Science and Technology, Zhejiang University

HIT-UltraVision

Title: Multi-level Wavelet Residual Network
Members:*Pengju Liu*[1,2]
(lpj008@126.com), Kai Zhang[1], Jingdong Liu[2], Jiye Liu[1], Hongzhi Zhang[1,2], Wangmeng Zuo[1]
Affiliation:
[1]School of Computer Science and Technology, 92 West Da Zhi St., Harbin Institute of Technology, Harbin, China
[2]Harbin Ultravision Technology, 6 Fuhua St., Harbin, China

Meteor

Title: Cascaded Residual Dense Network for Real Image Super-Resolution
Members:*Wenyi Tang*
(wenyitang@outlook.com)
Affiliation:
Individual

ECNU

Title: Encoder-Decoder architecture for Real Super-Resolution
Members:*Jing Liu*
(splinter02@163.com)
Affiliation:
East China Normal University(ECNU), Multimedia and Computer Vision Lab

AiDDle

Title: Stacked Lossless Deconvolutional Network
Members:*Youngjung Kim*
(read12300@add.re.kr), Changyeop Shin, Minbeom Kim, Sungho Kim
Affiliation:
Agency for Defence Development

BOE-IOT-AIBD

Title: BackProjection Pipelines
Members: *Pablo Navarrete Michelini*
(pnavarre@boe.com.cn), Hanwen Liu, Dan Zhu
Affiliation:
BOE Technology Group Co., Ltd.

xuxu123

Title: Spatial Color Attention Networks for Real Single Image Super-Resolution
Members: *Xuan Xu*
(xuxu@mix.wvu.edu), Xin Li
Affiliation:
West Virginia University

VIPSL

Title: Enhanced Residual Channel Attention Network for Image Super-resolution
Members: *Furui Bai*[1]
(frbai@stu.xidian.edu.cn), Xiaopeng Sun[1], Lin Zha[2], Yuanfei Huang[1], Wen Lu[1]
Affiliation:
[1]Xidian University
[2]Kiwi-image Technologies Co.,Ltd

zju231

Title: Orientation Attention Network for Real Image Super-Resolution
Members:*Yanpeng Cao*[1,2]
(caoyp@zju.edu.cn), Du Chen[1,2], Zewei He[1,2], Sun Anshun[1,2]; Siliang Tang[3]
Affiliation:
[1]State Key Laboratory of Fluid Power and Mechatronic Systems, School of Mechanical Engineering, Zhejiang University, Hangzhou, 310027, China
[2]Key Laboratory of Advanced Manufacturing Technology of Zhejiang Province, School of Mechanical Engineering, Zhejiang University, Hangzhou, 310027, China
[3]College of Computer Science and Technology, Zhejiang University, Hangzhou, 310027, China.

KSC

Title: Pyramid based Deep Residual Networks for Single Image Super-Resolution
Members:*Hongfei Fan*
(fanhongfei@kingsoft.com), Xiang Li, Guo Li, Wenjie Zhang, Yumei Zhang, Qingwen He
Affiliation:
Beijing Kingsoft Cloud Technology Co., Ltd, China

MENet

Title: Modified RCAN
Members:*Jinghui Qin*

(qinjingh@mail2.sysu.edu.cn), Lishan Huang, Yukai Shi, Pengxu Wei, Wushao Wen, Liang Lin
Affiliation:
Sun Yat-sen University

Early

Title: Multi-path block for efficient detail extraction
Members:*Jun Yu*
(xiegc@mail.ustc.edu.cn), Guochen Xie, Mengyan Li
Affiliation:
University of Science and Technology of China

Egg126

Title: Multiple Scale Residual with Spatial and Channel Attention Network for Super-Resolution
Members:*Lin Zha*
(LinZha@ymail.com), Furui Bai, XiaoPeng Sun, Wen Lu
Affiliation:
XiDian University, Kiwi-image Technologies Co.,ltd

XMU_IMAGE

Title: Learning the Degradation Process for Real-world Super-Resolution
Members:*Rong Chen*
(chenrong_mail@qq.com), Xiaotong Luo, Chen Hong, Yanyun Qu, Cuihua Li
Affiliation:
School of information science and engineering, Xiamen University

DeepSR

Title: Hierarchical Back Projection Network
Members:*Zhi-Song Liu*
(zhisong.liu@connect.polyu.hk), Li-Wen Wang, Chu-Tak Li
Affiliation:
Department of Electronic and Information Engineering, The Hong Kong Polytechnic University

Pikachu&Sparky&Co.

Title: Effects of Data Augmentation on Super Resolution
Members:*Can Zhao*
(volcanofly@gmail.com), Bowen Li
Affiliation:
Johns Hopkins University

QCAM

Title: Reduce latency for real-time single-image super resolution application significance
Members:*Chung-Chi Tsai*
(charles0184@gmail.com), Shang-Chih Chuang, Joon-Hee Choi, Joonsoo Kim, Xiaoyun Jiang
Affiliation:
Qualcomm Technology Inc.

Newbee

Title: Multi-scale gate fusion network
Members:*Ze Pan*
(panze0352@163.com), Qunbo Lv, Zheng Tan, Peidong He
Affiliation:
The Key Laboratory of computational optics, Chinese Academy of Science

References

[1] Y. Blau, R. Mechrez, R. Timofte, T. Michaeli, and L. Zelnik-Manor. The 2018 PIRM challenge on perceptual image super-resolution. In *ECCV*, 2018. 1

[2] J. Cai, H. Zeng, H. Yong, Z. Cao, and L. Zhang. Toward real-world single image super-resolution: A new benchmark and a new model. *arXiv preprint arXiv:1904.00523*, 2019. 2

[3] H. Chang, D.-Y. Yeung, and Y. Xiong. Super-resolution through neighbor embedding. In *CVPR*, 2004. 1

[4] D. Chen, Z. He, A. Sun, Y. Jiangxin, Y. Cao, Y. Cao, S. Tang, and M. Y. Yang. Orientation-aware deep neural network for real image super-resolution. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2019. 8

[5] Y. Chen, J. Li, H. Xiao, X. Jin, S. Yan, and J. Feng. Dual path networks. *CoRR*, abs/1707.01629, 2017. 4

[6] G. Cheng, A. Matsune, Q. Li, L. Zhu, H. Zang, and S. Zhan. Encoder-decoder residual network for real super-resolution. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2019. 6

[7] C. Dong, C. C. Loy, K. He, and X. Tang. Learning a deep convolutional network for image super-resolution. In *ECCV*, 2014. 1

[8] W. Dong, L. Zhang, G. Shi, and X. Wu. Image deblurring and super-resolution by adaptive sparse domain selection and adaptive regularization. *IEEE Transactions on Image Processing*, 20(7):1838–1857, 2011. 1

[9] R. Feng, J. Gu, Y. Qiao, and C. Dong. Suppressing model overfitting for image super-resolution networks. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2019. 4

[10] G. Freedman and R. Fattal. Image and video upscaling from local self-examples. *ACM Transactions on Graphics (TOG)*, 30(2):12, 2011. 1

[11] W. T. Freeman, E. C. Pasztor, and O. T. Carmichael. Learning low-level vision. *International journal of computer vision*, 40(1):25–47, 2000. 1

[12] S. Gao and X. Zhuang. Multi-scale deep neural networks for real image super-resolution. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2019. 3

[13] D. Glasner, S. Bagon, and M. Irani. Super-resolution from a single image. In *ICCV*, pages 349–356, 2009. 1

[14] J. Hu, L. Shen, and G. Sun. Squeeze-and-excitation networks. In *CVPR*, pages 7132–7141, 2018. 8

[15] J.-B. Huang, A. Singh, and N. Ahuja. Single image super-resolution from transformed self-exemplars. In *CVPR*. 1

[16] K. Jia, X. Wang, and X. Tang. Image transformation based on learning dictionaries across image spaces. *IEEE transactions on pattern analysis and machine intelligence*, 35(2):367–380, 2013. 1

[17] J. Johnson, A. Alahi, and L. Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In *ECCV*, 2016. 1

[18] S. R. K. He, X. Zhang and J. Sun. Deep residual learning for image recognition. Computer Vision and Pattern Recognition, 2016. 6

[19] J. Kim, J. Kwon Lee, and K. Mu Lee. Accurate image super-resolution using very deep convolutional networks. In *CVPR*, 2016. 1

[20] K. I. Kim and Y. Kwon. Single-image super-resolution using sparse regression and natural image prior. *IEEE transactions on pattern analysis and machine intelligence*, 32(6):1127–1133, 2010. 1

[21] T. Köhler, M. Bätz, F. Naderi, A. Kaup, A. Maier, and C. Riess. Bridging the simulated-to-real gap: Benchmarking super-resolution on real data. *arXiv preprint arXiv:1809.06420*, 2018. 2

[22] J. Kwak and D. Son. Fractal residual network and solutions for real super-resolution. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2019. 5

[23] Y. LeCun, Y. Bengio, and G. Hinton. Deep learning. *nature*, 521(7553):436, 2015. 1

[24] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, et al. Photo-realistic single image super-resolution using a generative adversarial network. In *CVPR*, 2017. 1, 9

[25] B. Lim, S. Son, H. Kim, S. Nah, and K. Mu Lee. Enhanced deep residual networks for single image super-resolution. In *CVPRW*, pages 136–144, 2017. 1

[26] Z.-S. Liu, L.-W. Wang, C.-T. Li, and W.-C. Siu. Hierarchical back projection network for image super-resolution. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2019. 9

[27] K. Z. L. L. P. Liu, H. Zhang and W. Zuo. Multi-level wavelet-cnn for image restoration. Computer Vision and Pattern Recognition Workshops, 2018. 6

[28] D. Park, J. Kim, and S. Y. Chun. Down-scaling with learned kernels in multi-scale deep neural networks for non-uniform single image deblurring. *arXiv preprint arXiv:1903.10157*, 2019. 4

[29] S. C. Park, M. K. Park, and M. G. Kang. Super-resolution image reconstruction: a technical overview. *IEEE signal processing magazine*, 20(3):21–36, 2003. 1

[30] M. S. Sajjadi, B. Scholkopf, and M. Hirsch. Enhancenet: Single image super-resolution through automated texture synthesis. In *ICCV*, 2017. 1

[31] S. Schulter, C. Leistner, and H. Bischof. Fast and accurate image upscaling with super-resolution forests. In *CVPR*, 2015. 1

[32] W. Shi, J. Caballero, F. Huszár, J. Totz, A. P. Aitken, R. Bishop, D. Rueckert, and Z. Wang. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In *CVPR*. 1, 8

[33] Y. Tai, J. Yang, and X. Liu. Image super-resolution via deep recursive residual network. In *CVPR*, 2017. 1

[34] R. Timofte, E. Agustsson, L. Van Gool, M.-H. Yang, and L. Zhang. Ntire 2017 challenge on single image super-resolution: Methods and results. In *CVPRW*, 2017. 1

[35] R. Timofte, V. De Smet, and L. Van Gool. Anchored neighborhood regression for fast example-based super-resolution. In *ICCV*, 2013. 1

[36] R. Timofte, S. Gu, J. Wu, and L. Van Gool. Ntire 2018 challenge on single image super-resolution: methods and results. In *CVPRW*, 2018. 1

[37] R. Timofte, R. Rothe, and L. 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. 2

[38] Z. Wang, A. C. Bovik, H. R. Sheikh, E. P. Simoncelli, et al. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004. 2

[39] Z. Wang, D. Liu, J. Yang, W. Han, and T. Huang. Deep networks for image super-resolution with sparse prior. In *ICCV*, 2015. 1

[40] X. Xu and X. Li. Scan: Spatial color attention networks for real single image super-resolution. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2019. 8

[41] C.-Y. Yang, C. Ma, and M.-H. Yang. Single-image super-resolution: A benchmark. In *ECCV*, 2014. 2

[42] J. Yang, Z. Lin, and S. Cohen. Fast image super-resolution based on in-place example regression. In *CVPR*. 1

[43] J. Yang, J. Wright, T. S. Huang, and Y. Ma. Image super-resolution via sparse representation. *IEEE transactions on image processing*, 19(11):2861–2873, 2010. 1

[44] W. Yang, X. Zhang, Y. Tian, W. Wang, and J.-H. Xue. Deep learning for single image super-resolution: A brief review. *arXiv preprint arXiv:1808.03344*, 2018. 1

[45] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu. Image super-resolution using very deep residual channel attention networks. In *ECCV*, 2018. 1, 2, 4, 5, 6, 7, 8, 9

[46] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu. Residual dense network for image super-resolution. In *CVPR*, 2018. 1