NTIRE 2019 Challenge on Video Super-Resolution: Methods and Results

Seungjun Nah Radu Timofte Shuhang Gu Sungyong Baik Seokil Hong
Gyeongsik Moon Sanghyun Son Kyoung Mu Lee Xintao Wang Kelvin C.K. Chan
Ke Yu Chao Dong Chen Change Loy Yuchen Fan Jiahui Yu Ding Liu
Thomas S. Huang Xiao Liu Chao Li Dongliang He Yukang Ding Shilei Wen
Fatih Porikli Ratheesh Kalarot Muhammad Haris Greg Shakhnarovich
Norimichi Ukita Peng Yi Zhongyuan Wang Kui Jiang Junjun Jiang Jiayi Ma
Hang Dong Xinyi Zhang Zhe Hu Kwanyoung Kim Dong Un Kang
Se Young Chun Kuldeep Purohit A. N. Rajagopalan Yapeng Tian Yulun Zhang
Yun Fu Chenliang Xu A. Murat Tekalp M. Akin Yilmaz Cansu Korkmaz
Manoj Sharma Megh Makwana Anuj Badhwar Ajay Pratap Singh
Avinash Upadhyay Rudrabha Mukhopadhyay Ankit Shukla Dheeraj Khanna
A. S. Mandal Santanu Chaudhury Si Miao Yongxin Zhu Xiao Huo

Abstract

This paper reviews the first NTIRE challenge on video super-resolution (restoration of rich details in low-resolution video frames) with focus on proposed solutions and results. A new REalistic and Diverse Scenes dataset (REDS) was employed. The challenge was divided into 2 tracks. Track 1 employed standard bicubic downscaling setup while Track 2 had realistic dynamic motion blurs. Each competition had 124 and 104 registered participants. There were total 14 teams in the final testing phase. They gauge the state-of-the-art in video super-resolution.

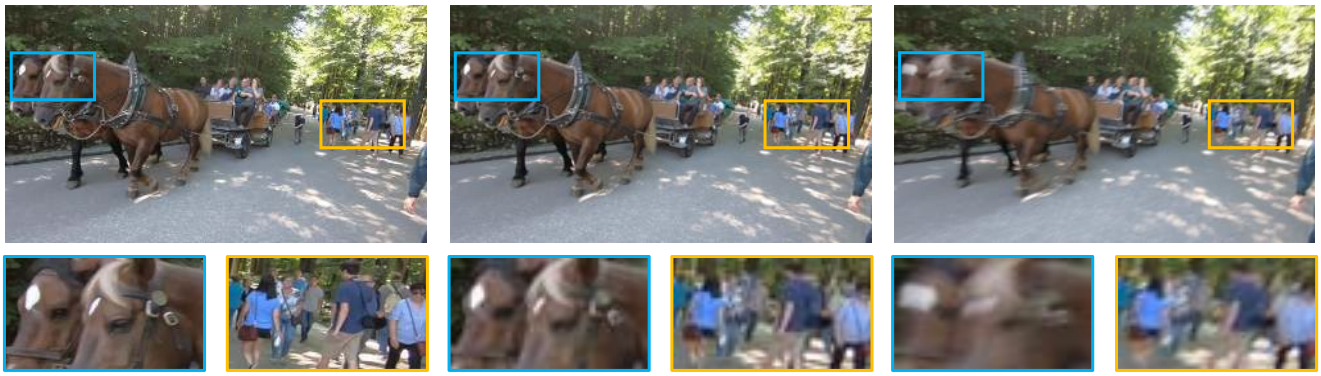
1. Introduction

Example-based video super-resolution (SR) aims at the restoration of the rich details (high frequencies) from low-resolution video frames based on a set of prior examples with low-resolution and high-resolution videos. The loss of contents can be caused by various factors such as quantization error, limitations of the sensor from the capturing camera, presence of defocus, motion blur or other degrading operators, and the use of downsampling operators to reduce the video resolution for storage purposes. Just like the conventional single image SR, video SR is also an ill-posed problem because for each low resolution (LR) frame, the space of corresponding high resolution (HR) frames can be very large.

In recent years, a significant amount of literature focused on video super-resolution research. The performance of the top methods continuously improved [28, 22, 15, 2, 24, 9, 21, 35], showing that the field is getting matured. However, when compared with single image super-resolution [1, 27, 29], video super-resolution lacks standardized benchmarks to allow for an assessment that is based on identical datasets and criteria. Recently, most of the video SR publications use the Vid4 [14] dataset for evaluation and comparison. Vid4 dataset contains 4 sequences and each video consists of 30 to 45 frames. The resolution of each frame is 480×704 or 576×704. In some works, other datasets are also proposed for evaluation like YT10 [21], Val4 [9], SPMCS [24], and CDVL [2]. However, they are not widely used for comparison yet. While those video super-resolution datasets have brought substantial improvements to this domain, they still have significant shortcomings: (1) they lack a standard training set: recent video SR works are trained from various sets that are chosen rather arbitrarily; (2) small test sets and resolution (often below HD resolution); (3) mixed downsampling methods (Gaussian blurs and bicubic kernels) that are not standardized; they are chosen for LR data generation and they are not consistent with single image SR literature where usually bicubic interpolation is employed.

The NTIRE 2019 video super-resolution challenge is a step forward in benchmarking and training of video super-resolution algorithms. It uses REalistic and Dynamic

S. Nah (seungjun.nah@gmail.com, Seoul National University), R. Timofte, S. Gu, S. Baik, S. Hong, G. Moon, S. Son and K.M. Lee 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



(a) Sharp HR (b) LR (Bicubic $\times$ 4 $\downarrow$) (c) LR (Bicubic $\times$ 4 $\downarrow$ + Blur)

Figure 1: Visualization of a video frame and its low resolution corresponding frames from the REDS dataset.

Scenes (REDS) dataset [16] consisting of 30000 reference frames with two types of degradation: the standard bicubic and additional dynamic motion blurs that are locally variant. Fig. 1 shows some images from REDS dataset. The REDS dataset is introduced in [16] along with a study of challenge results. In the next, we describe the challenge, present and discuss the results and describe the proposed methods.

2. NTIRE 2019 Challenge

The objectives of the NTIRE 2019 challenge on video super-resolution are: (i) to gauge and push the state-of-the-art in video super-resolution; (ii) to compare different solutions; (iii) to promote a novel large dataset (REDS); and (iv) to promote more challenging video super-resolution settings.

2.1. REDS Dataset

As a step forward from the previously proposed super-resolution and deblurring datasets, a novel dataset is promoted, namely REDS dataset [16]. It consists of 300 video sequences containing 100 frames of 720×1280 resolution. 240 sequences are for training, 30 for validation and the rest 30 for testing purposes. The frames are of high quality in terms of the reference frames, diverse scenes and locations, and realistic approximations of motion blur. REDS covers a large diversity of contents, people, handmade objects and environments (cities).

All the videos used to create the REDS dataset are manually recorded with GoPro HERO6 Black. They were originally recorded in 1920×1080 resolution at 120 fps. We calibrated the camera response function using [20] with regularization. Then the frames are interpolated [19] to virtually increase the frame rate up to 1920 fps so that averaged frames could exhibit smooth and realistic blurs without step artifacts. Then, the virtual frames are averaged in the signal space to mimic camera imaging pipeline [17]. To suppress noise, compression artifacts, we downscale reference sharp and corresponding blurry frames to 720×1280 resolution. This preprocessing also increases the effective number of pixels per resolution. The result blurry video frames resemble 24 fps video captured at duty cycle $\tau = 0.8$. Then, the sharp and blurry frames are $\times 4$ downscaled with the bicubic kernel to generate low-resolution videos.

2.2. Tracks and competitions

Track 1: Clean facilitates easy deployment of many video super-resolution methods. It assumes that the degradation only comes from downscaling. We generate each LR frame from the HR REDS frame by using MATLAB function `imresize` with bicubic interpolation and downscaling factor 4.

Track 2: Blur goes one step ahead and considers motion blur from fast-moving objects or shaken cameras as well. No Gaussian or other types of noise is added to the frames, but only motion blur from dynamic scenes is incorporated. We obtain each blurry LR frame following the procedure described in Section 2.1. More details are provided in [16]. The blur is locally variant and any further information such as blur strength or kernel shape was not provided. Each ground truth HR RGB frame from REDS is bicubically downscaled to the corresponding LR frames and used either for training, validation, or testing of the methods.

Competitions Both video deblurring challenge tracks are hosted as Codalab competitions. CodaLab platform was used for all of the NTIRE 2019 challenges competitions. Each participant is required to register to the CodaLab challenge tracks to access the data and submit their super-resolved results.

Challenge phases (1) Development (training) phase: the participants got both LR and HR train video frames and

the LR frames of the validation set. The participants had the opportunity to test their solutions on the LR validation frames and to receive feedback by uploading their results to the server. Due to the large-scale of the validation dataset, every 10^{th} frame was involved in evaluation. A validation leaderboard is available; (2) Final evaluation (test) phase: the participants got the sharp HR validation frames with the LR test frames. They had to submit both the super-resolved frames and a description of their methods before the challenge deadline. One week later, the final results were made available to the participants. The final results reflect the performance on every frame of the test set.

Evaluation protocol The Peak Signal-to-Noise Ratio (PSNR) measured in deciBels (dB) and the Structural Similarity Index (SSIM) [34] computed between a result frame and the ground truth are the quantitative measures. The higher the scores are the better the restoration fidelity to the ground truth frame. Because of boundary effects which may appear in particular methods, we ignore a rim of 1 pixel during the evaluation.

3. Challenge Results

From 124 and 104 registered participants for the competitions, 14 teams entered in the final phase and submitted results, codes, executables, and factsheets. Table 1 reports the final scoring results of the challenge and Table 2 shows the runtimes and the major details for each entry as provided by the authors in their factsheets. Section 4 describes the method of each team briefly while in the Appendix A are the team members and affiliations.

Use of temporal information All the proposed methods use the end-to-end deep learning and employ the GPU(s) for both training and testing. Interestingly, in contrast to the recent RNN-based video super-resolution methods, most teams (HelloVSR, UIUC-IFP, SuperRior, CyberverseSanDiego, XJTU-IAIR, BMIPL_UNIST, IPCV_IITM, Lucky Bird, mvgl) aggregated several video frames in channel dimension and let CNN learn the temporal relation to deblur a target frame. External optical flow estimation or warping was employed in none of the submitted methods. TTI used a recurrent model inspired from (DBPN [6]). CristianoRonaldo used a single image super-resolution method.

Restoration fidelity HelloVSR, UIUC-IFP, and SuperRior are the best scoring teams. HelloVSR is ***the winner*** of NTIRE 2019 Video Super-Resolution Challenge. HelloVSR achieves 31.79 dB for Track 1 and 30.17 dB for Track 2 improving +5.31 dB and +6.12 dB over the input low-resolution video, respectively. HelloVSR team achieves the best results for both of the competition tracks. Their solution shows consistent performance across the tracks and is also valid for NTIRE 2019 Video Deblurring Challenge [18].

Runtime / efficiency In Fig. 2 and 3, we plot the running time per image vs. achieved PSNR performance for both tracks. UIUC-IFP's solution showed good trade-off between the restoration quality in terms of PSNR and the running time. It runs in 0.98 s per frame for both Tracks on Tesla V100 in contrast to most other methods consume more than 1 seconds per frame. They had 0.71 dB gap with the HelloVSR's method in Track 2. Lucky Bird team's method was the fastest, taking only 0.013 seconds to process a frame.

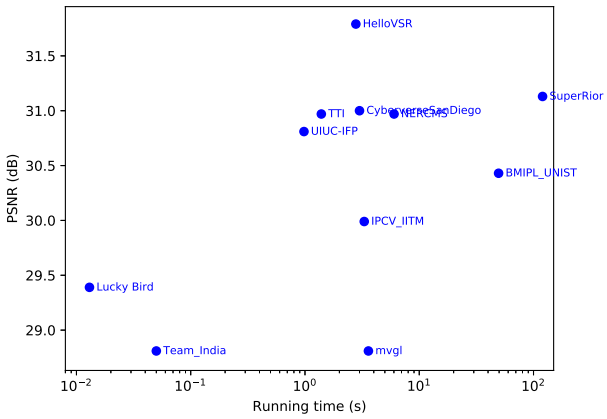


Figure 2: Runtime vs. performance for Track 1: Clean.

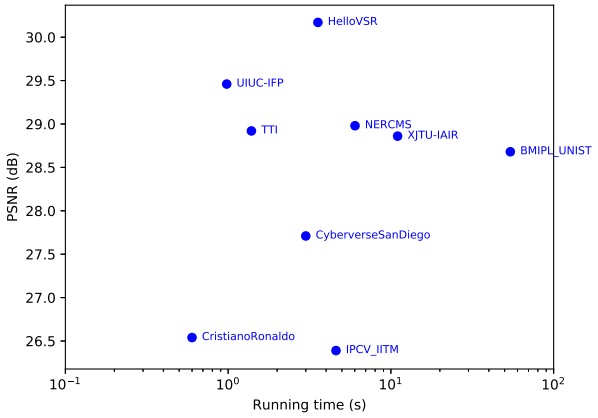


Figure 3: Runtime vs. performance for Track 2: Blur.

Ensembles Many solutions used self-ensemble [30] that averages the results from flipped and rotated inputs at test time. HelloVSR did not use rotation to reduce computation. SuperRior team focused on the fusion of multiple architectures. RDN [38], RCAN [37], DUF [9] are modified to take channel-concatenated frames as input and they esti-

Team	Author	Track 1: Clean PSNR	Track 1: Clean SSIM	Track 2: Blur PSNR	Track 2: Blur SSIM
HelloVSR	xixihaha	31.79 (1)	0.8962	30.17 (1)	0.8647
UIUC-IFP	fyc0624	30.81 (6)	0.8748	29.46 (2)	0.8430
SuperRior	lchkou	31.13 (2)	0.8811	-	-
CyberverseSanDiego	CyberverseSanDiego	31.00 (3)	0.8822	27.71 (7)	0.8067
TTI	iim_lab	30.97 (4)	0.8804	28.92 (4)	0.8333
NERCMS	Mrobot0	30.91 (5)	0.8782	28.98 (3)	0.8307
XJTU-IAIR	Hang	-	-	28.86 (5)	0.8301
BMIPL_UNIST	UNIST_BMIPL	30.43 (7)	0.8666	28.68 (6)	0.8252
IPCV_IITM	kuldeeppurohit3	29.99 (8)	0.8570	26.39 (9)	0.7699
Lucky Bird	NEU_SMILE_Lab	29.39 (9)	0.8419	-	-
mvgl	akinyilmaz	28.81 (10)	0.8249	-	-
Team_India	Manoj	28.81 (10)	0.8241	-	-
withdrawn team		28.54 (11)	0.8179	26.54 (8)	0.7587
CristianoRonaldo	ChristianoRonaldo	-	-	26.34 (10)	0.7549
Bicubic	*baseline*	26.48	0.7799	24.05	0.6809

Table 1: NTIRE 2019 Video Super-Resolution Challenge results on the REDS test data. HelloVSR team is the winner of the challenge with consistent performance in both tracks.

Team	Track 1 Clean	Track 2 Blur	Platform	GPU (at runtime)	Ensemble / Fusion (at runtime)
HelloVSR	2.788	3.562	PyTorch	TITAN Xp	Flip (x4)
UIUC-IFP	0.980	0.980	PyTorch	Tesla V100	Flip/Rotation (x8)
SuperRior	120.000	-	PyTorch	Tesla V100	Flip/Rotation/Temporal flip (x16) Adaptive model ensemble
CyberverseSanDiego	3.000	3.000	TensorFlow	RTX 2080 Ti	-
TTI	1.390	1.390	PyTorch	TITAN X	-
NERCMS	6.020	6.020	PyTorch	GTX 1080 Ti	Flip/Rotation (x8)
XJTU-IAIR	-	13.000	PyTorch	GTX 1080 Ti	Flip/Rotation (x8)
BMIPL_UNIST	45.300	54.200	PyTorch	TITAN V	-
IPCV_IITM	3.300	4.600	PyTorch	TITAN X	Flip/Rotation (x8)
Lucky Bird	0.013	-	PyTorch	TITAN Xp	-
mvgl	3.500	-	PyTorch	GTX 1080 Ti	-
Team_India	0.050	-	Pytorch/Tensorflow	Tesla V100	-
withdrawn team	398.000	398.000	-	-	-
CristianoRonaldo	-	0.600	TensorFlow	Tesla K80	-

Table 2: Reported runtimes per frame on REDS test data and details from the factsheets

mated the score maps for each output to generate spatially adaptive ensemble model. They also adopted temporal flips of inputs at test time for additional ensemble as well as spatial flips and rotations.

Train data REDS dataset [16] has 24000 train frames and all the participants found the amount of data to be sufficient for training their models. Training data augmentation strategy [30] such as flips and rotations by 90 degrees were employed by most of the participants.

Conclusions From the analysis of the presented results, we conclude that the proposed methods gauge the state-of-the-art performance in video super-resolution. The methods proposed by the best ranking team (HelloVSR) exhibit consistent superiority in both tracks in terms of PSNR and SSIM.

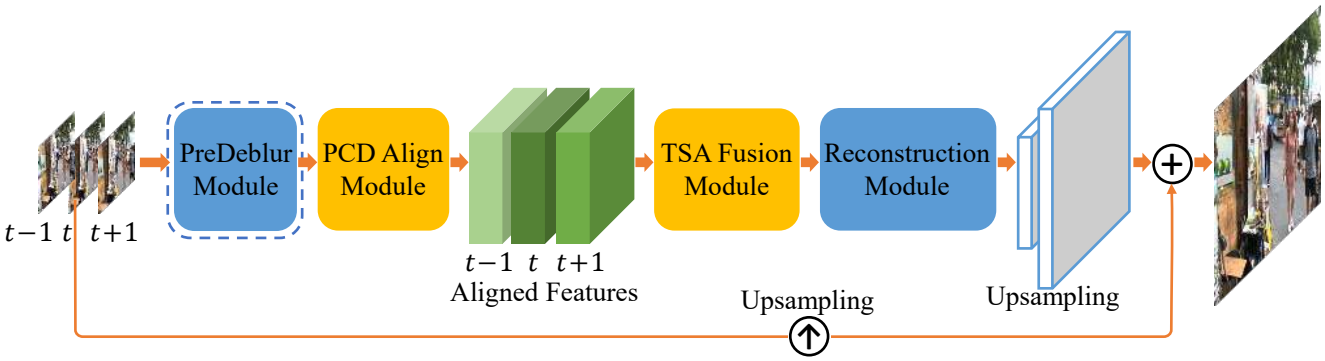


Figure 4: HelloVSR team: the proposed EDVR framework.

4. Challenge Methods and Teams

4.1. HelloVSR team

HelloVSR team proposes the EDVR framework [31], which takes $2N + 1$ low-resolution frames as inputs and generates a high-resolution output, as shown in Fig. 4. First, to alleviate the effects of blurry frames on alignment, a PreDeblur module is used to pre-process the blurry inputs before alignment (it is not included in the model for SR-clean track). Then, each neighboring frame is aligned to the reference frame by the PCD alignment module at the feature level. The TSA fusion module is used to fuse the aligned features effectively. The fused features then pass through a reconstruction module, which consists of several residual blocks [13] in EDVR and can be easily replaced by any other advanced modules in single image SR [11, 38, 6, 37, 33]. The upsampling operation is performed at the end of the network to increase the spatial size. Finally, the high-resolution reference frame is obtained by adding the predicted image residual to a direct upsampled image [10]. Note that EDVR is a generic architecture also suitable for other video restoration tasks, such as deblurring.

To address large and complex motions between frames, which are common in the REDS dataset, they propose a Pyramid, Cascading and Deformable convolution (PCD) alignment module. In this module, deformable convolutions [3, 26] is adopted to align frames at the feature level. They use a pyramid structure that first aligns features in lower scales with coarse estimations, and then propagates the offsets and aligned features to higher scales to facilitate precise motion compensation, similar to the notion adopted in optical flow estimation [8, 23]. Moreover, an additional deformable convolution is cascaded after the pyramidal alignment This approach further improve robustness of the alignment. The overview of the PCD module is shown in Fig. 5.

Since different frames and locations are not equally informative due to the imperfect alignment and imbalanced blur among frames, a Temporal and Spatial Attention (TSA) fusion module is designed to dynamically aggregate neighboring frames in pixel-level, as shown in Fig. 5. Temporal attention is introduced by computing the element-wise correlation between the reference frame and each neighboring frame in an embedding space. The correlation coefficients then weigh each adjacent feature at each location. Then, weighted features from all frames are convolved and fused together. After the fusion, they further apply spatial attention [35, 32, 37] to assign weights to each location in each channel to exploit cross-channel and spatial information more effectively.

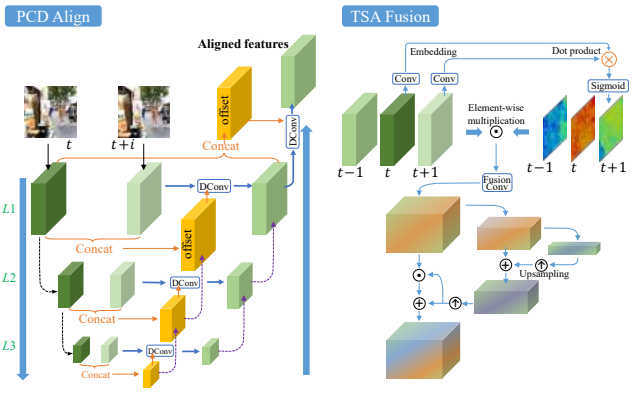


Figure 5: PCD alignment module and TSA fusion module in EDVR.

They also use a two-stage strategy to boost performance further. Specifically, a similar but shallower EDVR network is cascaded to refine the output frames of the first stage. The cascaded network can further remove the severe motion blur that cannot be handled by the preceding model and alleviate the inconsistency among output frames.

4.2. UIUC-IFP team

UIUC-IFP team proposes a new method, WDVR, which is based on WDSR [36, 5]. To achieve a better speed-accuracy trade-off, they investigate the intersection of three dimensions in deep video restoration networks: spatial, channel, and temporal. They enumerate various network architectures ranging from 2D convolutional models to 3D convolutional models and delve into their gains and losses in terms of training time, model size, boundary effects, prediction accuracy, and the visual quality of the restored videos. Under a strictly controlled computational budget, they explore the designs of each residual building block in a video restoration network, which consists of a mixture of 2D and 3D convolutional layers.

From their explorations, the summarized findings are: (1) In 3D convolutional models, setting more computation/channels for spatial convolution leads to better performance than on temporal convolution. (2) The best variant of 3D convolutional models is better than 2D convolutional models, but the performance gap is close. (3) In a very limited range, the performance can be improved by the increase of window size (5 frames for 2D model) or padding size (6 frames for 3D model). Based on these findings, they introduce the WDVR, wide-activated 3D convolutional network for video restoration, which achieves a better accuracy under similar computational budgets and runtime latency.

Multiple 3D wide-activated residual blocks in 3D models are designed with different ratio of parameters for temporal modeling. The most straightforward design of 3D is inflated version of 2D wide-Activated blocks, named as IAI. To reduce the ratio of temporal parameters by half, the inflated 3D convolution after activation is replaced with Spatial convolution, named as IAS. To further reduce the ratio, the other inflated 3D convolution is decomposed to Spatio-Temporal convolution, named as STAS. The spatio-temporal convolution explicitly isolates the parameters for spatial and temporal modelling. In STAS, the temporal convolution is connected with activations, so it has more channels than input and output features. Switching the order of Spatio-Temporal convolution, named as TSAS, can reduce temporal parameters even more, by moving temporal convolution to connect narrow block inputs.

4.3. SuperRior team

SuperRior team proposes to learn deep spatial-temporal features for up-sampling video frames by adapting multiple state-of-the-art image super-resolution methods [12] as shown in Fig. 6. They focus on the ensemble of different architectures that are independently designed, RDN [38], RCAN [37], DUF [9]. They were originally proposed for single-image super-resolution and the first layers are modified to accept concatenated frames. They estimate the spatial weight map for the output from each architecture and

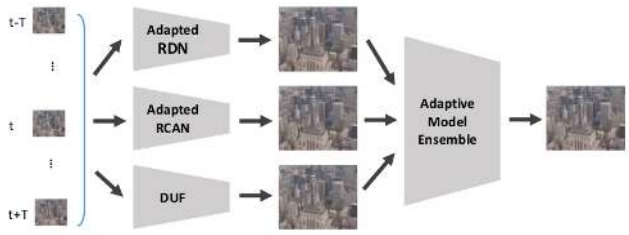


Figure 6: SuperRior team: proposed pipeline.

perform adaptive ensemble on them. At test time, temporal flips as well as spatial flips and rotations [30] are employed to further improve performance. Their adaptation schema can largely reduce the computation cost compared with using 3D based solutions.

4.4. CyberverseSanDiego team

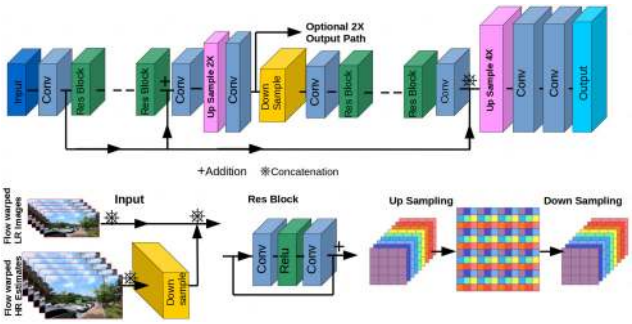


Figure 7: CyberverseSandiego team: overall architecture.

CyberverseSanDiego team proposes a fully convolutional neural network model for 4× video super-resolution that is capable of generating sharp video frames with high-resolution details by taking advantage of motion compensated reference frames and reusing the estimated high-resolution versions of a frame in previous stages for a bootstrapped (in other words, recurrent) resolution enhancement process.

They employ multiple motion-compensated reference frames of the current frame. To encourage temporally consistent results, they use a bootstrapped frame-recurrent approach where the reconstructed high-resolution frame of the previous step is dispatched into the network after rearranging its pixels into multiple low-resolution images. Their model in Fig. 7 consists of three main components; an input subnetwork that shuffles and combines multiple motion-compensated reference frames, a blending backbone that applies fully convolutional blocks on low-resolution feature maps, and a spatial upsampling subnetwork that reconstructs the high-resolution image.

4.5. TTI team

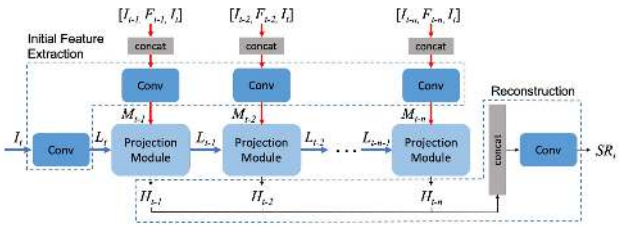


Figure 8: TTI team: proposed scheme.

TTI team proposes a novel architecture for the problem of video super-resolution as shown in Fig. 8. They integrate spatial and temporal contexts from consecutive video frames using a recurrent encoder-decoder module, that fuses multi-frame information with the more traditional, single frame super-resolution path for the target frame. In contrast to most prior work where frames are pooled together by stacking or warping, their model, the Recurrent Back-Projection Network (RBPN) [7] treats each context frame as a separate source of information. These sources are combined in an iterative refinement framework inspired by the idea of back-projection in multiple-image super-resolution. This is aided by explicitly representing estimated inter-frame motion with respect to the target, rather than explicitly aligning frames.

4.6. NERCMS team

NERCMS team proposes a progressive fusion video super-resolution, where the temporal correlations are extracted gradually. In particular, the method fuses multiple frames progressively and enlarges them at last, instead of fusing frames at first.

4.7. XJTU-IAIR team

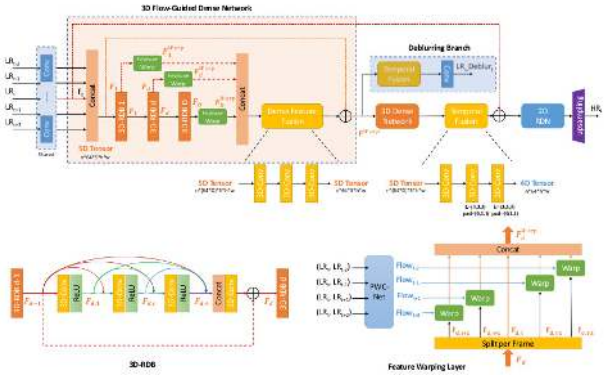


Figure 9: XJTU-IAIR team: proposed solution.

XJTU-IAIR team proposes a flow-guided spatio-temporal dense network (FSTDN) for the joint video deblurring and super-resolution task as shown in Fig. 9. The method estimates the optical flows among the consecutive frames and exploits the temporal information to reconstruct high-resolution images based on the estimated flow.

4.8. BMIPL_UNIST team

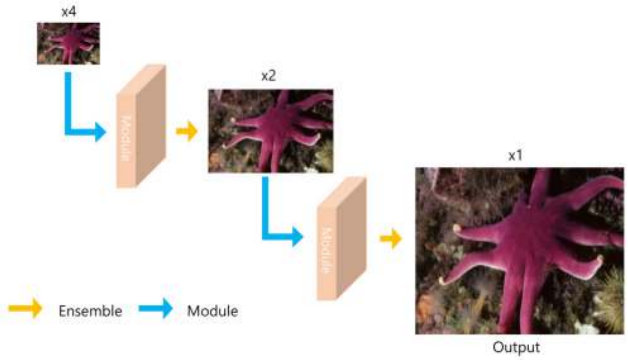


Figure 10: BMIPL_UNIST team: proposed pipeline.

BMIPL_UNIST team proposes a network with a module base as shown in Fig. 10. The baseline for each module is RCAN network for scale 2. The entire network consists of two modules. The output resolution of the first module is same as $\times 2$ downsampled frame. The output resolution of the second module is same as ground-truth frame. In each module, the network takes three sequential images such as $t-1$, t, $t+1$-th frames as input. In each SR model, they set the number of res-group as 10 and the number of res-block as 20. For training, they firstly train 1st module and train 2nd module by using 1st module pre-trained weight. The author used ensemble method in each module to generate the training data.

4.9. IPCV_IITM team

IPCV_IITM team feeds one single frame, as well as 5 neighbor frames to two branches of our network, and at the end, they fuse both the output to get the final super-resolved frame as shown in Fig. 11. For single image super-resolution branch, they employ ESRGAN [33] architecture for extracting HR information from each frame. For multi-frame information integration for super-resolution, they train a compact EDSR [13] type architecture. They also utilize the pre-trained FlowNet [4] to align input LR frames during training and testing. To handle motion blurs in Track 2, they additionally use deblurring stage afterward, feeding the output of the super-resolution model to the deblurring model.

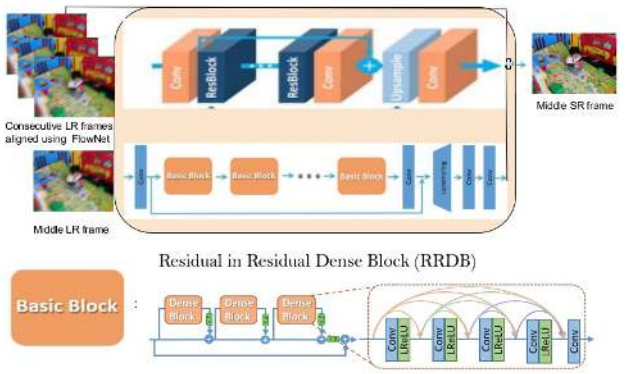


Figure 11: IPCV_IITM team: proposed solution.

4.10. Lucky Bird team

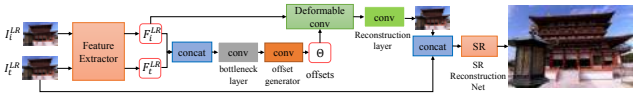


Figure 12: Lucky Bird team: overall pipeline of the method.

Lucky Bird team proposes a temporal deformable alignment network (TDAN) to adaptively align the reference frame and each supporting frame at the feature level without computing optical flow as shown in Fig. 12. The TDAN uses features from both the reference frame and each supporting frame to dynamically predict offsets of sampling convolution kernels. By using the corresponding kernels, TDAN transforms supporting frames to align with the reference frame. The author utilized a reconstruction network which takes aligned frames and the reference frame to predict HR video frames.

4.11. mvgl team

mvgl team proposes a model that consists of two separate fully convolutional networks. The first network called FCTNN takes 9 consecutive frames to produce the super-resolved version of the middle frame. Inspired by [13], they use residual network structure as a deep architecture. For upsampling, they take the sub-pixel layer which contains convolution and pixel-shuffle operations. Since FCTNN super-resolves the middle frame of 9 consecutive frames, they lose the first and last 4 frames in a sequence. To produce these 8 frames at test time, they designed the second architecture which handles single image super-resolution.

4.12. Team_India team

Team_India team proposes a deep back projection network [6] based model at frame level to spatially upsample the frames then quality is further improved by using RED-10 to remove artifacts which arises during upsampling of frames.

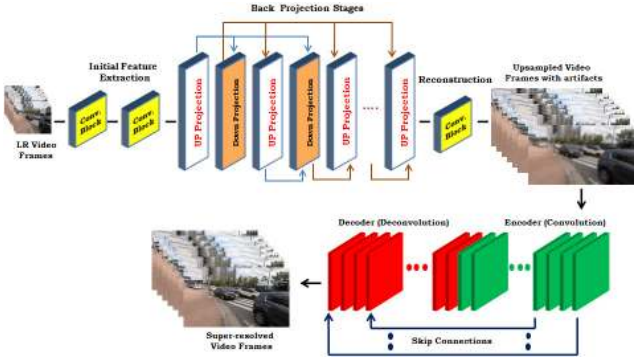


Figure 13: Team_India team: overall pipeline of the model.

4.13. CristianoRonaldo team

CristianoRonaldo team proposes a network that is jointly trained for both image deblurring and super-resolution. SRN-DeblurNet [25] is used for pretraining deblur module. Owing to the proposed network and joint training method, the model can generate three scales of sharp images($\times 1$, $\times 2$, $\times 4$). The method does not require consecutive neighbor frames to super-resolve a target frame. Temporal information is not necessary, and the method can be applied to a single image as well.

Acknowledgments

We thank the NTIRE 2019 sponsors: OPPO Mobile Corp., Ltd., NVIDIA Corp., HUAWEI Technologies Co. Ltd., SAMSUNG Electronics Co., Ltd., Amazon.com, Inc., MediaTek Inc., and ETH Zurich.

A. Teams and affiliations

NTIRE2019 team

Title: NTIRE 2019 Challenge on Video Super-Resolution
Members: *Seungjun Nah*[1] *(seungjun.nah@gmail.com)*, Radu Timofte[2], Shuhang Gu[2], Sungyong Baik[1], Seokil Hong[1], Gyeongsik Moon[1], Sanghyun Son[1], Kyoung Mu Lee[1]
Affiliations:
[1] Department of ECE, ASRI, SNU, Korea
[2] Computer Vision Lab, ETH Zurich, Switzerland

HelloVSR

Title: EDVR: Video Restoration with Enhanced Deformable Convolutional Networks

Members: *Xintao Wang*[1] *(xintao.alpha@gmail.com)*, Kelvin C.K. Chan[2], Ke Yu[1], Chao Dong[3], Chen Change Loy[2]
Affiliations:
[1] The Chinese University of Hong Kong, Hong Kong
[2] Nanyang Technological University, Singapore
[3] Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, China

UIUC-IFP

Title: Wide Activation for Efficient and Accurate Video Super-Resolution
Members: *Yuchen Fan (yuchenf4@illinois.edu)*, Jiahui Yu, Ding Liu, Thomas S. Huang
Affiliation:
University of Illinois at Urbana-Champaign, US

SuperRior

Title: Adapting Image Super-Resolution State-of-the-arts and Learning Adaptive Multi-Model Ensemble for Video Super-Resolution
Members: *Xiao Liu (liuxiao12@baidu.com)*, Chao Li, Dongliang He, Yukang Ding, Shilei Wen
Affiliation:
Computer Vision Technology (VIS) Department, Baidu Inc., China

CyberverseSanDiego

Title: RecNet: Recursive Network for Clean Video Super-Resolution
Members: *Fatih Porikli (fatih.porikli@huawei.com)*, Ratheesh Kalarot
Affiliation:
O-Lab, San Diego Device and Hardware, US

TTI

Title: Recurrent Back-Projection Networks
Members: *Muhammad Haris*[1] *(mharis@toyota-ti.ac.jp)*, Greg Shakhnarovich[2], Norimichi Ukita[1]
Affiliations:
[1] Toyota Technological Institute (TTI), Japan
[2] Toyota Technological Institute at Chicago (TTIC), US

NERCMS

Title: Progressive Fusion Video Super-Resolution
Members: *Peng Yi (yipeng@whu.edu.cn)*, Zhongyuan Wang, Kui Jiang, Junjun Jiang, Jiayi Ma
Affiliation:
National Engineering Research Center for Multimedia Software, China

XJTU-IAIR

Title: Flow-Guided Dense Spatio-Temporal Network for Joint Video Deblurring and Super-Resolution
Members: *Hang Dong*[1] *(dhunter1230@gmail.com)*, Xinyi Zhang[1], Zhe Hu[2]
Affiliations:
[1] National Engineering Laboratory for Vision Information Processing and Application, Xian Jiaotong University, China
[2] Hikvision Research

BMIPL_UNIST

Title: Efficient Module based Video Super-Resolution for Multiple Problems
Members: *Kwanyoung Kim (cubeyoung@unist.ac.kr)*, Dong Un Kang, Se Young Chun
Affiliation:
Ulsan National Institute of Science and Technology (UNIST), Korea

IPCV_IITM

Title: Multi-Branch Motion-compensated Network for Video Super-Resolution
Members: *Kuldeep Purohit (kuldeeppurohit3@gmail.com)*, A. N. Rajagopalan
Affiliation:
Indian Institute of Technology Madras, India

Lucky Bird

Title: Temporally-Deformable Alignment Network for Video Super-Resolution
Members: *Yapeng Tian*[1] *(yapengtian@rochester.edu)*, *Yulun Zhang*[2] *(yulun100@gmail.com)*, Yun Fu[2], Chenliang Xu[1]
Affiliations:
[1] Department of CS, University of Rochester, US
[2] Department of ECE, Northeastern University, US

mvgl

Title: Fully Convolutional Temporal Neural Network (FCTNN) for Video Super-Resolution
Members: *A. Murat Tekalp (mtekalp@ku.edu.tr)*, M. Akin Yilmaz, Cansu Korkmaz
Affiliation:
Koc University, Turkey

Team_India

Title: Improved Video Super-Resolution using Enhanced Deep Back Projection Network
Members: *Manoj Sharma (mksnith@gmail.com)*, Megh Makwana, Anuj Badhwar, Ajay Pratap Singh, Avinash Upadhyay, Rudrabha Mukhopadhyay, Ankit Shukla, Dheeraj Khanna, A. S. Mandal, Santanu Chaudhury
Affiliations:
CEERI-Delhi Center, IIT Delhi, CCS Computers Pvt Ltd., Delhi Technological University, India

CristianoRonaldo

Title: Joint Single Image Deblurring and Super-Resolution
Members: *Si Miao (miaosi2018@sari.ac.cn)*, Yongxin Zhu, Xian Huo
Affiliation:
Shanghai Advanced Research Institute, Chinese Academy of Sciences, China

References

[1] 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. 1

[2] Jose Caballero, Christian Ledig, Andrew Aitken, Alejandro Acosta, Johannes Totz, Zehan Wang, and Wenzhe Shi. Real-time video super-resolution with spatio-temporal networks and motion compensation. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017. 1

[3] Jifeng Dai, Haozhi Qi, Yuwen Xiong, Yi Li, Guodong Zhang, Han Hu, and Yichen Wei. Deformable convolutional networks. In *The IEEE International Conference on Computer Vision (ICCV)*, Oct 2017. 5

[4] Alexey Dosovitskiy, Philipp Fischer, Eddy Ilg, Philip Hausser, Caner Hazirbas, Vladimir Golkov, Patrick van der Smagt, Daniel Cremers, and Thomas Brox. FlowNet: Learning optical flow with convolutional networks. In *The IEEE International Conference on Computer Vision (ICCV)*, December 2015. 7

[5] Yuchen Fan, Jiahui Yu, and Thomas S. Huang. Wide-activated deep residual networks based restoration for bpg-compressed images. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2018. 6

[6] Muhammad Haris, Gregory Shakhnarovich, and Norimichi Ukita. Deep back-projection networks for super-resolution. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018. 3, 5, 8

[7] Muhammad Haris, Greg Shakhnarovich, and Norimichi Ukita. Recurrent back-projection network for video super-resolution. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019. 7

[8] Tak-Wai Hui, Xiaoou Tang, and Chen Change Loy. Liteflownet: A lightweight convolutional neural network for optical flow estimation. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018. 5

[9] Younghyun Jo, Seoung Wug Oh, Jaeyeon Kang, and Seon Joo Kim. Deep video super-resolution network using dynamic upsampling filters without explicit motion compensation. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018. 1, 3, 6

[10] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee. Accurate image super-resolution using very deep convolutional networks. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016. 5

[11] Christian Ledig, Lucas Theis, Ferenc Huszar, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, and Wenzhe Shi. Photo-realistic single image super-resolution using a generative adversarial network. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017. 5

[12] Chao Li, Dongliang He, Xiao Liu, Yukang Ding, and Shilei Wen. Adapting image super-resolution state-of-the-arts and learning multi-model ensemble for video super-resolution. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2019. 6

[13] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. Enhanced deep residual networks for single image super-resolution. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, July 2017. 5, 7, 8

[14] Ce Liu and Deqing Sun. A bayesian approach to adaptive video super resolution. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2011. 1

[15] Ding Liu, Zhaowen Wang, Yuchen Fan, Xianming Liu, Zhangyang Wang, Shiyu Chang, and Thomas Huang. Robust video super-resolution with learned temporal dynamics. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2507–2515, 2017. 1

[16] Seungjun Nah, Sungyong Baik, Seokil Hong, Gyeongsik Moon, Sanghyun Son, Radu Timofte, and Kyoung Mu Lee. Ntire 2019 challenges on video deblurring and super-resolution: Dataset and study. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2019. 2, 4

[17] Seungjun Nah, Tae Hyun Kim, and Kyoung Mu Lee. Deep multi-scale convolutional neural network for dynamic scene deblurring. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017. 2

[18] Seungjun Nah, Radu Timofte, Sungyong Baik, Seokil Hong, Gyeongsik Moon, Sanghyun Son, Kyoung Mu Lee, Xintao Wang, Kelvin C.K. Chan, Ke Yu, Chao Dong, Chen Change Loy, Yuchen Fan, Jiahui Yu, Ding Liu, Thomas S. Huang, Hyeonjun Sim, Munchurl Kim, Dongwon Park, Jisoo Kim, Se Young Chun, Muhammad Haris, Greg Shakhnarovich, Norimichi Ukita, Syed Waqas Zamir, Aditya Arora, Salman Khan, Fahad Shahbaz Khan, Ling Shao, Rahul Kumar Gupta, Vishal Chudasama, Heena Patel, Kishor Upla,

Hongfei Fan, Guo Li, Yumei Zhang, Xiang Li, Wenjie Zhang, Qingwen He, Kuldeep Purohit, A. N. Rajagopalan, Jeonghun Kim, Mohammad Tofighi, Tiantong Guo, and Vishal Monga. Ntire 2019 challenge on video deblurring: Methods and results. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2019. 3

[19] Simon Niklaus, Long Mai, and Feng Liu. Video frame interpolation via adaptive separable convolution. In *IEEE International Conference on Computer Vision (ICCV)*, Oct 2017. 2

[20] Mark A. Robertson, Sean Borman, and Robert L. Stevenson. Dynamic range improvement through multiple exposures. In *The IEEE International Conference on Image Processing (ICIP)*, Oct 1999. 2

[21] Mehdi S. M. Sajjadi, Raviteja Vemulapalli, and Matthew Brown. Frame-recurrent video super-resolution. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018. 1

[22] Wenzhe Shi, Jose Caballero, Ferenc Huszar, 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 *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016. 1

[23] Deqing Sun, Xiaodong Yang, Ming-Yu Liu, and Jan Kautz. PWC-Net: CNNs for optical flow using pyramid, warping, and cost volume. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018. 5

[24] Xin Tao, Hongyun Gao, Renjie Liao, Jue Wang, and Jiaya Jia. Detail-revealing deep video super-resolution. In *The IEEE International Conference on Computer Vision (ICCV)*, Oct 2017. 1

[25] Xin Tao, Hongyun Gao, Xiaoyong Shen, Jue Wang, and Jiaya Jia. Scale-recurrent network for deep image deblurring. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018. 8

[26] Yapeng Tian, Yulun Zhang, Yun Fu, and Chenliang Xu. TDAN: temporally deformable alignment network for video super-resolution. *CoRR*, abs/1812.02898, 2018. 5

[27] Radu Timofte, Eirikur Agustsson, Luc Van Gool, Ming-Hsuan Yang, and Lei Zhang. 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

[28] Radu Timofte, Vincent De Smet, and Luc Van Gool. A+: Adjusted anchored neighborhood regression for fast super-resolution. In *Asian Conference on Computer Vision*, pages 111–126. Springer, 2014. 1

[29] 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

[30] 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, 4, 6

[31] Xintao Wang, Kelvin C.K. Chan, Ke Yu, Chao Dong, and Chen Change Loy. Edvr: Video restoration with enhanced deformable convolutional networks. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2019. 5

[32] Xintao Wang, Ke Yu, Chao Dong, and Chen Change Loy. Recovering realistic texture in image super-resolution by deep spatial feature transform. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018. 5

[33] 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 *The European Conference on Computer Vision (ECCV) Workshops*, September 2018. 5, 7

[34] Zhou Wang, Alan C Bovik, Hamid R Sheikh, Eero P Simoncelli, et al. Image quality assessment: from error visibility to structural similarity. *The IEEE Transactions on Image Processing (TIP)*, 13(4):600–612, 2004. 3

[35] Sanghyun Woo, Jongchan Park, Joon-Young Lee, and In So Kweon. CBAM: convolutional block attention module. In *The European Conference on Computer Vision (ECCV)*, September 2018. 1, 5

[36] Jiahui Yu, Yuchen Fan, Jianchao Yang, Ning Xu, Zhaowen Wang, Xinchao Wang, and Thomas S. Huang. Wide activation for efficient and accurate image super-resolution. *CoRR*, abs/1808.08718, 2018. 6

[37] Yulun Zhang, Kunpeng Li, Kai Li, Lichen Wang, Bineng Zhong, and Yun Fu. Image super-resolution using very deep residual channel attention networks. In *The European Conference on Computer Vision (ECCV)*, September 2018. 3, 5, 6

[38] Yulun Zhang, Yapeng Tian, Yu Kong, Bineng Zhong, and Yun Fu. Residual dense network for image super-resolution. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018. 3, 5, 6