

NTIRE 2019 Challenge on Video Deblurring and Super-Resolution: Dataset and Study

Seungjun Nah¹ Sungyong Baik¹ Seokil Hong¹ Gyeongsik Moon¹ Sanghyun Son¹
 Radu Timofte² Kyoung Mu Lee¹

¹Department of ECE, ASRI, SNU, Korea ²CVL, ETH Zurich & Merantix GmbH
 seungjun.nah@gmail.com, {dsybaik, hongceo96, mks0601, thstkdgus35}@snu.ac.kr,
 radu.timofte@vision.ee.ethz.ch, kyoungmu@snu.ac.kr

Abstract

This paper introduces a novel large dataset for video deblurring, video super-resolution and studies the state-of-the-art as emerged from the NTIRE 2019 video restoration challenges. The video deblurring and video super-resolution challenges are each the first challenge of its kind, with 4 competitions, hundreds of participants and tens of proposed solutions. Our newly collected *REalistic and Di-verse Scenes dataset (REDS)* was employed by the challenges. In our study, we compare the solutions from the challenges to a set of representative methods from the literature and evaluate them on our proposed REDS dataset. We find that the NTIRE 2019 challenges push the state-of-the-art in video deblurring and super-resolution, reaching compelling performance on our newly proposed REDS dataset.

1. Introduction

Example-based video deblurring and super-resolution aim to recover the rich details and the clear texture from blurry and low-resolution video frames, based on prior examples under the form of degraded blurry and low-resolution (LR) and corresponding sharp and high-resolution (HR) videos. The loss of contents can be caused by various factors such as quantization error, limitations of the sensor from the capturing camera, shakes from hand-held cameras, fast moving objects, compression at saving time, or other degrading operators and the use of downscaling operators to reduce the video resolution for storage purposes. Video deblurring and video super-resolution are representative ill-posed problems in visual quality restoration problems as the space of the corresponding sharp HR video is very large.

Video deblurring [37, 12, 28, 35, 13, 15] as well as video super-resolution [27, 22, 3, 29, 9, 26, 36] have received

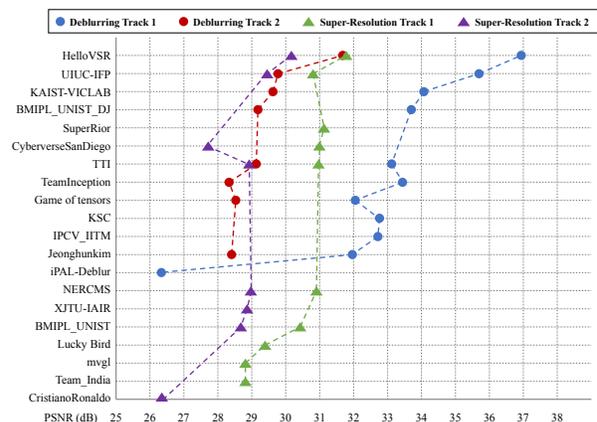


Figure 1: Representative methods from the challenges and their average PSNR performance on REDS dataset for each competition track.

much interest in the literature of research. Still, as seen in the single image super-resolution literature [30, 8, 32, 1, 31, 2, 4] for further progress in video deblurring and SR, standardized benchmarks are essential to allow comparison of different proposed methods under the same conditions.

In recent video deblurring works, [23, 28, 35] datasets that synthesize realistic motion blurs are popular for training and evaluation. However, there are different pros and cons in the blur synthesis techniques for each dataset. For video super-resolution, Vid4 dataset [21] with 155 frames is commonly used for comparison and each work employs different training datasets [26, 9, 29, 3].

In this work, we propose a novel REDS dataset with *REalistic and Dynamic Scenes* of 720×1280 resolution high-quality video frames collected by ourselves. It has 30000 frames with various contents, locations, natural and hand-made objects. Moreover, we organized the first example-



Figure 2: Visualization of the proposed REDS validation and test frames. REDS contains 240, 30, 30 sequences for training, validation, test, respectively. Each sequence has 100 frame length.

based video deblurring and video super-resolution online challenges which used the REDS dataset. The NTIRE 2019 challenges employ 4 types of degradations and corresponding competition tracks: motion blur, motion blur with compression artifacts, bicubic downscaling, and bicubic downscaling with motion blur. The degradation information is not given and only known through the training data pairs of degraded and ground truth frames.

Another contribution of this paper is a study of our newly proposed REDS with the achieved performance by the winners of the NTIRE 2019 video deblurring and video super-resolution challenges and representative methods from recent years. We report results using a selection of image quality measures for benchmarking. Fig. 1 shows submitted solutions for the NTIRE 2019 challenge and their achievements evaluated on the REDS dataset.

The remainder of the paper is structured as follows. Section 2 introduces the REDS dataset. Section 3 reviews the NTIRE 2019 video challenges and its settings. Section 4 introduces the image quality assessment (IQA) measures, Section 5 - the datasets, and Section 6 - the methods from our study. Section 7 concludes the paper.

2. Proposed REDS Dataset

We propose the REDS dataset, a novel REAListic and Dynamic Scenes dataset of 720×1280 resolution for training and benchmarking example-based video deblurring and super-resolution (see Fig. 2). REDS is intended to complement the existing video deblurring and SR datasets (see

Fig. 7) to increase the content diversity and provide more realism in degradation, especially, motion blur.

Recording: We manually recorded 300 RGB video clips, paying attention to the quality of each frame, diversity of source contents (scenes and locations) and dynamics of various motion. We used the GoPro HERO6 Black camera to record videos of 1080×1920 resolution at 120 fps. In contrast to the previous datasets for deblurring that captured videos in higher frame rate (240 fps) [23, 28], we choose slower frame rate for better quality. Note that most consumer-level high-speed cameras don't access all of the cell arrays during the readout time. Under the limited computational power of the camera processors, decreasing the frame rate allows access to more sensor array elements, increasing the number of effective pixels per frame. Each frame remains sharp when the shutter speed is fast. Still, the effective pixels are less than the full resolution. Also, noise or MPEG lossy compression could bring some artifacts.

Frame interpolation: Motion blur occurs due to the dynamics during the camera exposure. Averaging the high-frame-rate video frames approximates the photograph taken at a longer exposure [23]. When the frame rate is not high enough, simply averaging frames may generate unnatural spikes or steps in the blur trajectory [35], especially when the resolution is high and the motion is fast. To fill in the missing information between the frames, we employed a CNN trained to interpolate frames [24]. We chose a learned CNN instead of optical flow to handle nonlinear motions

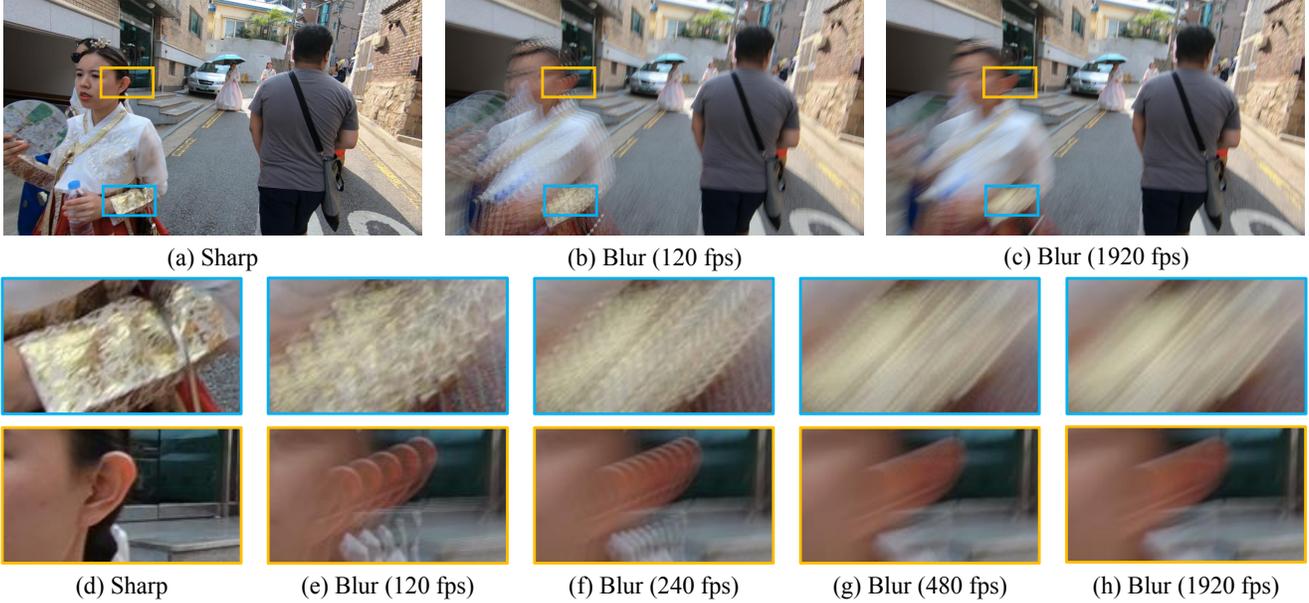


Figure 3: Visual comparison of the synthesized blur by the virtual frame rate of video. Averaging frames at 120 or 240 fps could cause unnatural blurs with artifacts in case of large motion. The noted fps refers to the virtual frame rate of interpolated videos that are averaged to create blurry frames.

and the warping artifacts. We increase the frame rate to virtual 1920 fps by recursively interpolating the frames. The effect of interpolation in the quality of blur is shown in Fig. 3.

Calibration: When taking a picture, the sensor signal is converted to RGB pixels by a nonlinear camera response function (CRF). We calibrated the inverse CRF using Robertson et al. [25] images captured at various exposures. As 8-bit representation saturates at value 255, the calibration could be inaccurate at higher pixel values ($p > 250$) when the calibration images are over-exposed. Hence, we replace the inverse CRF at $p > 250$ by appending a linear function having a slope of the inverse CRF at $p = 250$. Here, p denotes the RGB pixel value. We visualize the estimated CRF in Fig. 4, compare with linear [28] and gamma function [23] that are assumed for blur synthesis.

Blur synthesis: We average the 1920 fps video frames to produce virtual 24 fps blurry video with duty cycle $\tau = 0.8$. The averaging is done in the signal space to mimic a camera imaging pipeline, using the estimated CRF and the inverse CRF. To further increase the per-pixel quality of the data, we suppress the noise and artifacts by downscaling both the synthesized blurry frames and the recorded sharp frames by $2/3$ to 720×1280 resolution. We used OpenCV function `resize` bicubic interpolation as it produces visually sharper results than MATLAB due to different parameter value. There are 300 sequences in total, and each sequence contains 100 pairs of the blurry and sharp frame. We use those generated blurry videos as input for the NTIRE 2019

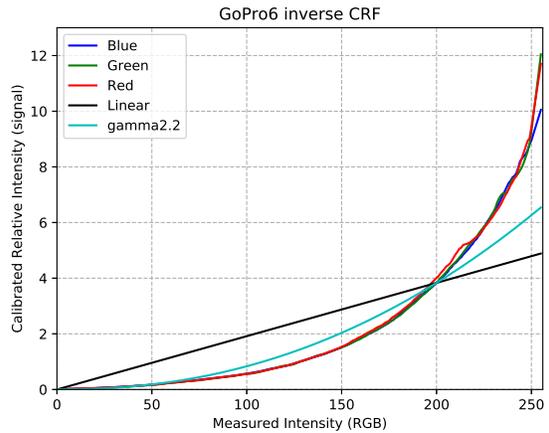


Figure 4: Calibrated inverse camera response function of GoPro HERO6 Black for RGB channels. It differs from the linear or gamma function assumptions from the previous datasets.

Video Deblurring Challenge Track 1: Clean.

Video compression: The above process was done to produce high-quality videos and blurs without realistic artifacts such as noise and compression. To promote the development of deblurring methods that apply to more realistic and common degradation, we compress the frames by saving the videos in mp4 (MPEG-4 Part 14) format. We used MATLAB `VideoWriter` to save the videos at 60% quality. Those compressed blurry videos are introduced to the

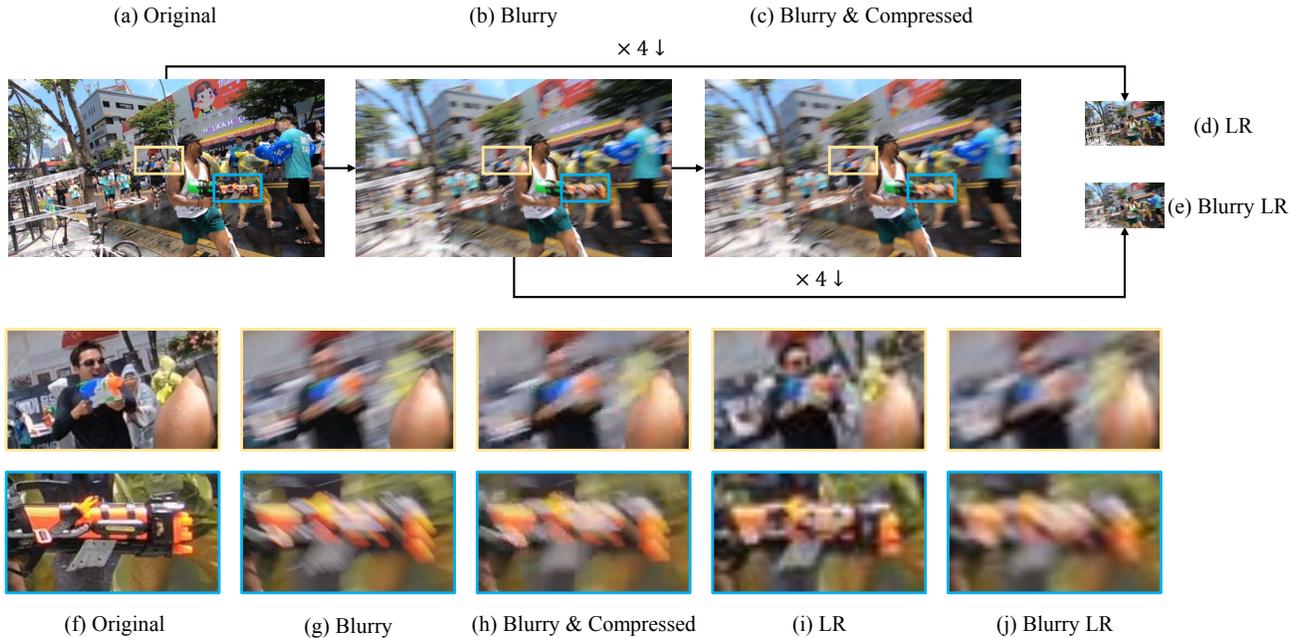


Figure 5: Visualization of the REDS dataset and the degradation types. In the video deblurring challenge, motion blurs (Track 1) and video compressed (Track 2) data were provided. The video super-resolution challenge provided the low-resolution of sharp (Track 1) and blurry (Track 2) frames.

NTIRE 2019 Video Deblurring Challenge: Track 2 Compression artifacts.

Downscaling: We also downscale the sharp and the blurry frames, respectively, to promote the development of example-based video super-resolution algorithms. They are employed in the NTIRE 2019 Video Super-Resolution Challenge: Track 1 Clean and Track 2 Blur. We used MATLAB function `imresize` bicubic interpolation at scale 4.

Diversity: We visited various countries, cities and towns, institutes and facilities, theme parks, festivals, palaces and castles, tourist attractions, historical places, zoos, stores, water parks, etc. to capture diverse scenes and objects. The contents include people from various nationalities, crowds, handmade objects, buildings, structures, artworks, furniture, vehicles, colorful textured clothes, and many other objects of different categories.

Partitions: After collecting and processing the REDS 300 video sequences, we computed the PSNR between the blurry and sharp frames. We split the REDS 300 sequences of frames into the train, validation, test sets. We randomly generated partitions of 240 train, 30 validation, and 30 test sequences until we achieved a good balance in quality. Fig. 2 visualizes part of the 30 sequences for validation and testing of the REDS dataset.

3. NTIRE 2019 Video Challenges

The NTIRE 2019 challenges on example-based video deblurring and super-resolution were the first of their kind and had the following objectives: to gauge the state-of-the-art in video deblurring and video super-resolution, to facilitate comparison of different solutions on a novel large dataset - REDS, and to propose more challenging and realistic deblurring and super-resolution settings. Fig. 5 shows an example of set of degraded images provided in each challenge (*i.e.*, deblurring and super-resolution).

Video Deblurring Challenge The challenge has two tracks: Track 1 for blurs (‘clean’) and Track 2 for additional MPEG compression (‘compression artifacts’). For Track 1, the degradation is a carefully processed realistic motion blur and facilitates easy deployment of recent solutions. Track 2 is more challenging as it uses a combination of blur and lossy compression. For both of the tracks, blur kernel is unknown. The compression method and the ratio were also unknown to the challenge participants. But they are implicitly known through exemplars of the blurry compressed frames and the corresponding sharp, uncompressed frames.

Video Super-Resolution Challenges The challenge has two tracks: Track 1 for bicubic downscaling (‘clean’) and Track 2 for blurs followed by bicubic downscaling (‘blur’).

For Track 1, the resizing is the popular bicubic downscaling by scale 4 and enables the straightforward application of the previous solutions. Track 2 is more challenging as the loss of information suffers from both of the downscaling and the motion blurs that are different for every frame and every pixel.

The hosting platform for the competitions is CodaLab¹.

For each competition, we provide the degraded (blurry, compressed, LR) and the reference (sharp HR) train frames (from the REDS train set) for training during the development phase. The phase allowed the participants to test their solutions on the validation frames (from REDS validation set) and compare their scores through an online validation server and associated leaderboard. Due to the massive size of the dataset, only every 10th frame was evaluated (300 frames). The final testing (evaluation) phase provided the degraded test frames (from REDS test set) and invited the submission of the restored results before the challenge deadline. In the testing phase, the full test set (3000 frames) was involved in measuring the scores and rankings. PSNR and SSIM are the challenge main quantitative quality assessment measures for the restoration results. Also, we ignore a 1-pixel image boundary from each image to minimize distortions from the boundary effect.

Challenge results Each competition had on average 100 registered participants and 13 ~ 14 teams submitted the final results, code/executables, and factsheets for the final test phase. All these competing solutions and the achieved results on the REDS test data are described in the NTIRE 2019 video deblurring and super-resolution challenge reports. All the proposed challenge solutions employ deep learning of convolutional neural networks and use GPUs for both training and testing. They propose a diversity of ideas and design details and generally build upon and go beyond the recent video deblurring [28, 13, 35, 15] and super-resolution [11, 3, 29] works. In Fig. 6, we plot the average PSNR vs. runtime results of the challenge solutions in comparison with several other representative methods. In Table 1, we show results for a selection of them. The challenge winning solutions are mostly consistent across the 4 competitions, indicating that the proposed solutions for the video deblurring and video super-resolution generalize well to each other. Also, PSNR and SSIM scores correlate well. The scores on Track 2 are generally worse than on Track 1 for the same methods/solutions for both of the tasks and reflects the increasing difficulty from the combination of degradation types.

4. Image Quality Assessment (IQA)

There are significant interests in the automatic assessment of the image quality, and many approaches have been

Method	Video Deblurring				Video Super-Resolution			
	Track 1		Track 2		Track 1		Track 2	
HelloVSR	36.96	0.9657	31.69	0.8783	31.79	0.8962	30.17	0.8649
UIUC-IFP	35.71	0.9522	29.78	0.8285	30.81	0.8748	29.46	0.8430
KAIST-VICLAB	34.09	0.9361	29.63	0.8261	-	-	-	-
BMPL_UNIST_DJ	33.71	0.9363	29.19	0.8190	-	-	-	-
SuperRior	-	-	-	-	31.13	0.8811	-	-
CyberverseSanDiego	-	-	-	-	31.00	0.8822	27.71	0.8067
no processing/bicubic	26.13	0.7749	25.40	0.7336	26.48	0.7799	24.05	0.6809

Table 1: Quantitative results on the REDS test set for 4 video restoration competitions.

proposed. With the presence of carefully generated dataset with ground truth reference frames, we mainly focus on the full reference measures.

When we have a ground truth image \mathbf{G} with C color channels and $H \times W$ pixels, the quality of a corresponding (whether degraded or restored) image I can be defined as the pixel-level fidelity to the ground truth. One of the most popular metrics are **Mean Square Error (MSE)** defined as Eq. 1. Another popular measure which is directly related to MSE is **Peak Signal-to-Noise Ratio (PSNR)** defined as Eq. 2. However, since minimizing MSE is equivalent to predicting a mean of possible solutions, MSE-based restoration models reconstruct blurry output images. Also, they are vulnerable to a simple translation, too.

$$\text{MSE} = \frac{1}{CHW} \sum_{c,h,w} (\mathbf{G}_{chw} - \mathbf{I}_{chw})^2 \quad (1)$$

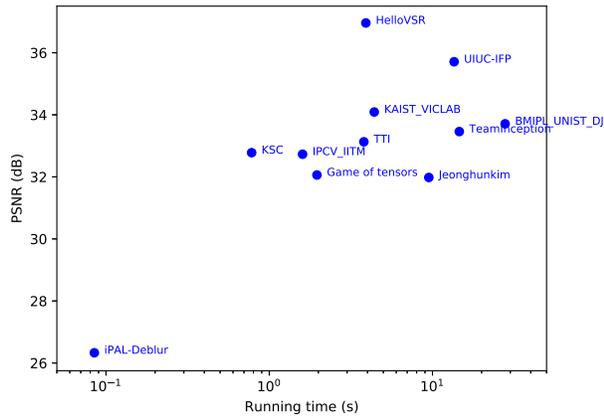
$$\text{PSNR} = 10 \log_{10} \frac{255^2}{\text{MSE}} \quad (2)$$

Another group of referenced metrics evaluates the image similarity in terms of the structure rather than the raw value. While MSE and PSNR measure the amount of error, the **Structural Similarity index (SSIM)** [34] is a perceptual quality based model that considers the degradation of images as changes in the perceived structural information.

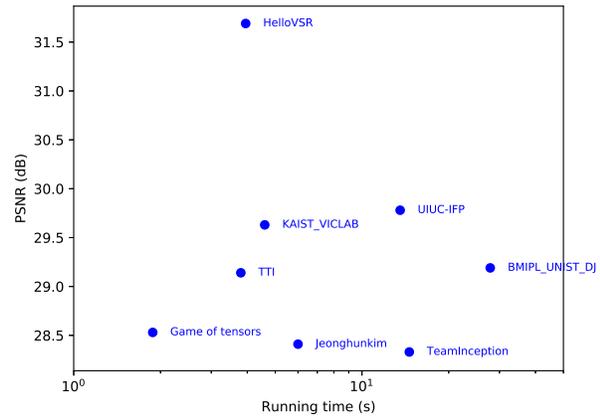
Above metrics are not designed to measure the quality of the restored images or videos from blur, compression, or low-resolution. However, they tend to generalize well for various types of image distortions as well and typically used to evaluate the accuracy of many methods that try to improve the visual quality.

As deblurring and super-resolution aim to recover the lost information such as detailed textures and high-frequency components from the latent image, an ideal IQA measure would use and reflect fidelity to the ground truth when it is available. However, often the ground truth is not available for real data, and the space of possible solutions is large. Therefore, plausible and perceptually qualitative restoration results are desirable as long as the information from the degraded images or videos is preserved. There have been several studies about deblurring and super-resolution which aimed to improve perceptual quality with adversarial and perceptual losses [23, 17, 18, 10, 39].

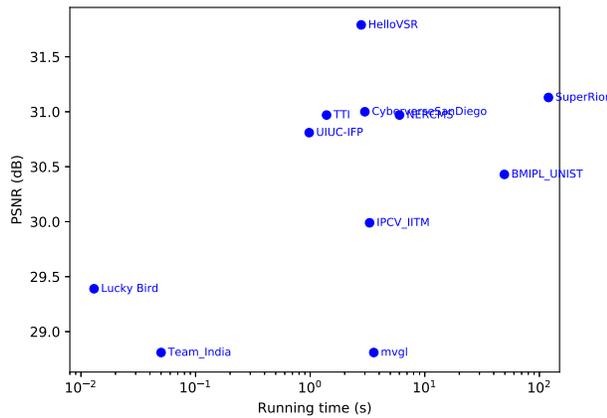
¹<https://competitions.codalab.org/>



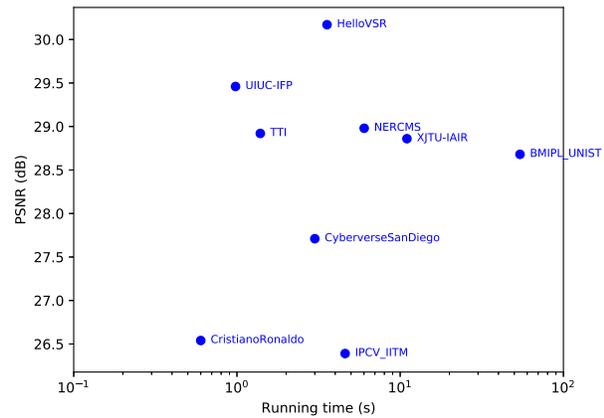
(a) Video Deblurring, Track 1



(b) Video Deblurring, Track 2



(c) Video Super-Resolution, Track 1



(d) Video Super-Resolution, Track 2

Figure 6: Runtime vs. PSNR results on NTIRE 2019 Video Deblurring & Super-Resolution Challenges on REDS test data.

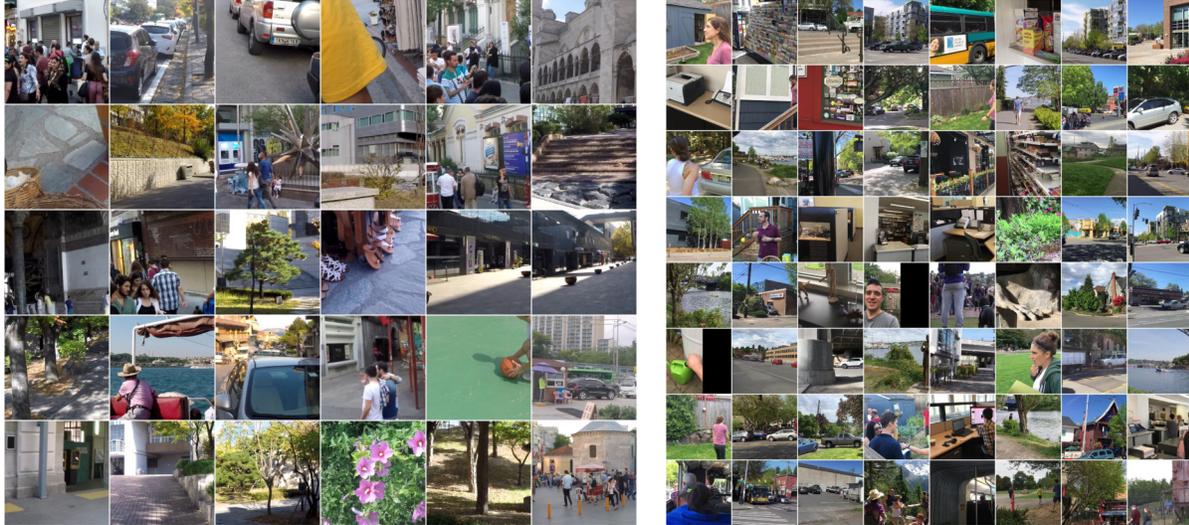
Most of the image/video restoration methods evaluate their performance on either the luminance component or full RGB channels. Luminance component (Y channel from the YCbCr color representation) is considered to be more important since the human perception typically recognizes the texture by the luminance while the changes in chroma component are less sensitive to the human eyes. In this challenge, we measure the quality metrics using the RGB channels to put more weight on the vividness and color as well as the luminance. We ignore the 1-pixel width boundary of the image from the evaluation.

5. Datasets

In this section, we describe the datasets that are studied in video deblurring and super-resolution literature for training and evaluation. Fig. 7 shows sample frames from commonly used video deblurring and super-resolution datasets.

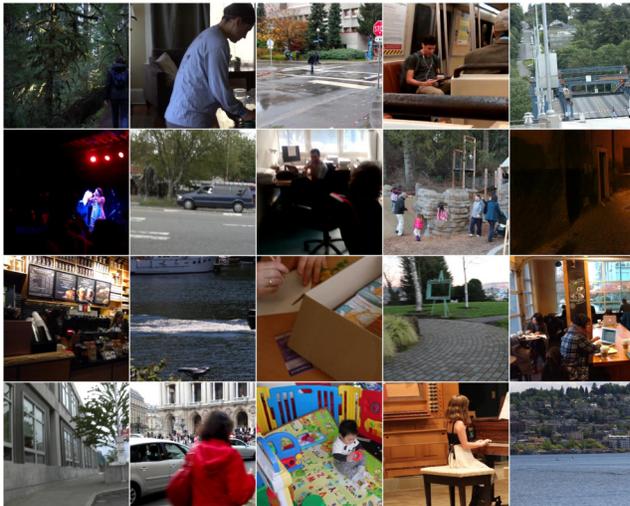
5.1. Video Deblurring Datasets

In the early studies of video deblurring, researchers used real blurry videos for their experiments [5, 12]. As the ground-truth sharp videos were not available, the perceptual quality was the primary way to compare different methods. Wulff and Black [37] presented a bilayered blur model that could have different blur status in the front and back layer segments. Köhler et al. [16] recorded the 6D motion of the camera and played it back to capture blurry and corresponding sharp images. To simulate such a blurring process in more general dynamic environments, Kim et al. [14] used a high-speed camera to take an average of sharp frames to synthesize blurs in HD (720×1280) resolution. Nah et al. [23] extended their data and presented GOPRO dataset consisting of 2103 training and 1111 testing frame pairs, assuming gamma function as CRF. Su et al. [28] used multiple cameras to present dataset containing 5708 training



(a) GOPRO dataset [23]

(b) DVD dataset [28]



(c) Real blurry videos [5]



(d) Vid4 dataset [21]

Figure 7: Visualization of the standard popular video datasets. (a), (b), and (c) are for deblurring. (d) is for super-resolution.

and 1000 testing frames. Wieschollek et al. [35] collected high-resolution videos from the web, interpolated frames with linear optical flow and downsampled them to generate smoother blurs for training.

5.2. Video Super-Resolution Datasets

Vid4 [21] is a commonly used dataset for evaluating video super-resolution methods. Vid4 contains a total of 155 frames from 4 sequences. The resolution of each frame varies 480×704 to 576×704 . However, those HR frames contain several artifacts from compression and noise. Vimeo-90k [38] is a recently proposed large-scale dataset containing 64612 training samples with 448×256 resolution. Also, several works evaluated their methods

on customized datasets to validate the proposed methods on higher resolution videos with rich details. Caballero et al. [3] used the CDVL database, training their model with 3000 FHD (1080×1920) frames. Tao et al. [29] proposed SPMCS dataset of FHD resolution videos. 945 sequences are for training, and the rest 30 sequences are for evaluation. Sajjadi et al. [26] also collected FHD videos from YouTube for evaluation (YT10). Jo et al. [9] used 4 videos as their validation set (Val4). These diverse training and evaluation datasets make it difficult to compare the solutions fairly. Furthermore, the downscaling methods are mixed up between Gaussian blurs and bicubic interpolation with different parameters.

5.3. REDS

Our proposed REDS dataset is intended to be in high-quality in terms of the reference frame and the realism of the degradation. We focus on making smooth and natural blurs while the compression artifact follows the standard codec and the downscaling is consistent with the single-image super-resolution literature. The dynamics between neighbor frames make the problem more challenging and promote the development of restoration methods. Refer to Section 2 for the dataset acquisition details.

6. NTIRE 2019 Video Deblurring and Super-Resolution Challenge Methods

In this study, we use the top methods from NTIRE 2019 Video Deblurring and Super-Resolution Challenges as well as several representative methods from recent literature.

6.1. NTIRE 2019 Video Challenge Methods

HelloVSR is the winner of the both NTIRE 2019 video deblurring and super-resolution challenges. They proposed the EDVR framework [33]. The consecutive frames go through PCD module to be aligned features to the target frame. Then TSA fusion module is used to fuse the temporal information between the features. Finally, the reconstruction module based on residual blocks [20] and upsampling module predicts the restored frame. Most of the operations are performed in a coarse scale as the deblurring model has downsampling layers in the front.

UIUC-IFP is a winner of NTIRE 2019 video deblurring and super-resolution challenges. They built, WDVR architecture inspired by WDSR [40, 7]. They investigate the effectiveness of 2D and 3D convolutions under a limited computational budget. The deep structure with 3D convolutional layers explores spatial and temporal context information jointly. In contrast to 2D convolution models where the temporal connection lies in the early channel fusion, the 3D convolution extracts the temporal relation gradually.

KAIST-VICLAB proposed a video deblurring model that has three parallel branches. One branch predicts motion *deblurring* kernel that is convolved on the input target frame to recover low-frequency components. Another branch generates the RGB residual image directly. They are linearly combined by the weight map which is produced by the other branch.

SuperRior team employed an adaptive ensemble model [19] for video super-resolution. They trained an adaptive ensemble module that generates a spatial weight map that averages different model that learns a spatial weight map for 3 different architected super-resolution models. They trained an ensemble model that generates a spatial weight map for RDN [42], RCAN [41], and DUF [9]. Each module architecture is modified to take

multi-frame inputs.

6.2. Other Representative Methods

We select several recently proposed methods for video deblurring and super-resolution and introduce them.

Video deblurring methods

DBN of Su et al. [28] applies channel-wise concatenation to multiple frames. Neighbor frames are warped to be aligned with a center frame using homography or optical flow. The encoder-decoder architecture fuses information from the aligned frames to deblur the center frame.

RDN of Wieschollek et al. [35] builds a recurrent network that reuses part of the features from the previous frame in multiple scales.

OVD of Kim et al. [13] employs a temporal blending module in a fast RNN architecture. The model learns blending parameters for injecting information from previous time steps to the current frame.

STTN of Kim et al. [15] introduces a spatiotemporal flow estimation module that captures long-range temporal dependencies. The module can also be applied in video super-resolution.

Video super-resolution methods

VSRNet of Kappeler et al. [11] presents CNN architectures with early fusion with concatenated input frames or extracted features from the frames. They enforce their convolutional kernels to be symmetric to accelerate training. Also, optical flow is applied to compensate for the motion between neighbor frames.

ESPCN of Caballero et al. [3] investigates into early fusion, slow fusion, and 3D convolutions to learn temporal relation. They propose to use a multi-scale spatial transformer for motion compensation. Sub-pixel convolution that makes computation efficient is employed instead of bicubic interpolation preprocessing, similarly to FSRCNN [6].

SPMC of Tao et al. [29] proposes a sub-pixel motion compensation layer for frame alignment with convolutional LSTM architecture.

7. Conclusion

In this paper, we introduced REDS, a new dataset for video deblurring and super-resolution benchmarking. We provide high-quality ground truth reference frames as well as corresponding degraded frames. Each degraded frame models commonly occurring video degradations such as motion blur, compression, and downsampling. We studied the winning solutions from the NTIRE 2019 video deblurring and super-resolution challenges in comparison with representative methods from the recent literature.

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.

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] Yochai Blau, Roey Mechrez, Radu Timofte, Tomer Michaeli, and Lihi Zelnik-Manor. The 2018 pirm challenge on perceptual image super-resolution. In *The European Conference on Computer Vision (ECCV) Workshops*, September 2018. 1
- [3] 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, 5, 7, 8
- [4] Jianrui Cai, Shuhang Gu, Radu Timofte, Lei Zhang, et al. Ntire 2019 challenge on real image super-resolution: Methods and results. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2019. 1
- [5] Sunghyun Cho, Jue Wang, and Seungyong Lee. Video deblurring for hand-held cameras using patch-based synthesis. *ACM Transactions on Graphics (TOG)*, 31(4):64, 2012. 6, 7
- [6] Chao Dong, Chen Change Loy, and Xiaoou Tang. Accelerating the super-resolution convolutional neural network. In *European conference on computer vision*, pages 391–407. Springer, 2016. 8
- [7] 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. 8
- [8] Jia-Bin Huang, Abhishek Singh, and Narendra Ahuja. Single image super-resolution from transformed self-exemplars. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5197–5206, 2015. 1
- [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, 7, 8
- [10] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In *The European Conference on Computer Vision (ECCV)*, October 2016. 5
- [11] Armin Kappeler, Seunghwan Yoo, Qiqin Dai, and Aggelos K Katsaggelos. Video super-resolution with convolutional neural networks. *IEEE Transactions on Computational Imaging*, 2(2):109–122, 2016. 5, 8
- [12] Tae Hyun Kim and Kyoung Mu Lee. Generalized video deblurring for dynamic scenes. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015. 1, 6
- [13] Tae Hyun Kim, Kyoung Mu Lee, Bernhard Scholkopf, and Michael Hirsch. Online video deblurring via dynamic temporal blending network. In *The IEEE International Conference on Computer Vision (ICCV)*, October 2017. 1, 5, 8
- [14] Tae Hyun Kim, Seungjun Nah, and Kyoung Mu Lee. Dynamic video deblurring using a locally adaptive blur model. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 40(10):2374–2387, 2018. 6
- [15] Tae Hyun Kim, Mehdi S. M. Sajjadi, Michael Hirsch, and Bernhard Scholkopf. Spatio-temporal transformer network for video restoration. In *The European Conference on Computer Vision (ECCV)*, September 2018. 1, 5, 8
- [16] Rolf Köhler, Michael Hirsch, Betty Mohler, Bernhard Schölkopf, and Stefan Harmeling. Recording and playback of camera shake: Benchmarking blind deconvolution with a real-world database. In *The European Conference on Computer Vision (ECCV)*, October 2012. 6
- [17] Orest Kupyn, Volodymyr Budzan, Mykola Mykhailych, Dmytro Mishkin, and Ji Matas. DeblurGAN: Blind motion deblurring using conditional adversarial networks. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018. 5
- [18] 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
- [19] 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. 8
- [20] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. Enhanced deep residual networks for single image super-resolution. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, July 2017. 8
- [21] 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, 7
- [22] Ding Liu, Zhaowen Wang, Yuchen Fan, Xianming Liu, Zhangyang Wang, Shiyu Chang, and Thomas Huang. Robust video super-resolution with learned temporal dynamics. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017. 1
- [23] 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. 1, 2, 3, 5, 6, 7
- [24] 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

- [25] Mark A. Robertson, Sean Borman, and Robert L. Stevenson. Dynamic range improvement through multiple exposures. In *The IEEE International Conference on Image Processing (ICIP)*, October 1999. 3
- [26] 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, 7
- [27] 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
- [28] Shuochen Su, Mauricio Delbracio, Jue Wang, Guillermo Sapiro, Wolfgang Heidrich, and Oliver Wang. Deep video deblurring for hand-held cameras. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017. 1, 2, 3, 5, 6, 7, 8
- [29] Xin Tao, Hongyun Gao, Renjie Liao, Jue Wang, and Jiaya Jia. Detail-revealing deep video super-resolution. In *IEEE International Conference on Computer Vision (ICCV)*, October 2017. 1, 5, 7, 8
- [30] Radu Timofte, Vincent De Smet, and Luc Van Gool. Anchored neighborhood regression for fast example-based super-resolution. In *The IEEE International Conference on Computer Vision (ICCV)*, December 2013. 1
- [31] 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
- [32] 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. 1
- [33] 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. 8
- [34] Zhou Wang, Alan C Bovik, Hamid R Sheikh, Eero P Simoncelli, et al. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing (TIP)*, 13(4):600–612, 2004. 5
- [35] Patrick Wieschollek, Michael Hirsch, Bernhard Scholkopf, and Hendrik P. A. Lensch. Learning blind motion deblurring. In *IEEE International Conference on Computer Vision (ICCV)*, Oct 2017. 1, 2, 5, 7, 8
- [36] 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
- [37] Jonas Wulff and Michael Julian Black. Modeling blurred video with layers. In *The European Conference on Computer Vision (ECCV)*, September 2014. 1, 6
- [38] Tianfan Xue, Jiajun Wu, Katherine L. Bouman, and William T. Freeman. Visual dynamics: Probabilistic future frame synthesis via cross convolutional networks. *CoRR*, abs/1607.02586, 2016. 7
- [39] Chih-Yuan Yang, Chao Ma, and Ming-Hsuan Yang. Single-image super-resolution: A benchmark. In *The European Conference on Computer Vision (ECCV)*, September 2014. 5
- [40] 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. 8
- [41] 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. 8
- [42] 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. 8