

NTIRE 2020 Challenge on Real-World Image Super-Resolution: Methods and Results

Andreas Lugmayr* Martin Danelljan* Radu Timofte* Namhyuk Ahn
 Dongwoon Bai Jie Cai Yun Cao Junyang Chen Kaihua Cheng SeYoung Chun
 Wei Deng Mostafa El-Khamy Chiu Man Ho Xiaozhong Ji Amin Kheradmand
 Gwantae Kim Hanseok Ko Kanghyu Lee Jungwon Lee Hao Li Ziluan Liu
 Zhi-Song Liu Shuai Liu Yunhua Lu Zibo Meng Pablo Navarrete Michelini
 Christian Micheloni Kalpesh Prajapati Haoyu Ren Yong Hyeok Seo Wan-Chi Siu
 Kyung-Ah Sohn Ying Tai Rao Muhammad Umer Shuangquan Wang
 Huibing Wang Timothy Haoning Wu Haoning Wu Biao Yang Fuzhi Yang
 Jaejun Yoo Tongtong Zhao Yuanbo Zhou Haijie Zhuo Ziyao Zong Xueyi Zou

Abstract

This paper reviews the NTIRE 2020 challenge on real world super-resolution. It focuses on the participating methods and final results. The challenge addresses the real world setting, where paired true high and low-resolution images are unavailable. For training, only one set of source input images is therefore provided along with a set of unpaired high-quality target images. In Track 1: Image Processing artifacts, the aim is to super-resolve images with synthetically generated image processing artifacts. This allows for quantitative benchmarking of the approaches w.r.t. a ground-truth image. In Track 2: Smartphone Images, real low-quality smart phone images have to be super-resolved. In both tracks, the ultimate goal is to achieve the best perceptual quality, evaluated using a human study. This is the second challenge on the subject, following AIM 2019, targeting to advance the state-of-the-art in super-resolution. To measure the performance we use the benchmark protocol from AIM 2019. In total 22 teams competed in the final testing phase, demonstrating new and innovative solutions to the problem.

1. Introduction

Single image Super-Resolution (SR) is the task of increasing the resolution of a given image by filling in additional high-frequency content. It has been a popular re-

*Andreas Lugmayr (andreas.lugmayr@vision.ee.ethz.ch), Martin Danelljan, and Radu Timofte at ETH Zürich are the NTIRE 2020 challenge organizers. The other authors participated in the challenge.

Appendix A contains the authors' team names and affiliations.
<https://data.vision.ee.ethz.ch/cvl/ntire20/>



Figure 1. Visual example of the input LR images and ground truth HR images used in the challenge. For Track 1 the input is generated with a common image signal processing operation to simulate the real-world SR case where we can measure against a undisclosed ground truth. For Track 2 we the input are untouched iPhone3 images. Both tracks have the goal to super-resolve to a clean target domain.

search topic for decades [33, 17, 56, 65, 63, 69, 70, 71, 61, 13, 28, 64, 14, 15, 38, 40, 43, 16, 2, 3, 26, 29, 21] due to its many applications. The current trend addresses the ill-posed SR problem using deep Convolutional Neural Networks (CNNs). While initial methods focused on achieving high fidelity in terms of PSNR [14, 15, 38, 40, 43]. Recent work has put further emphasis on generating perceptually more appealing predictions using for instance adversarial losses [72, 41, 67].

Deep learning based SR methods are known to consume large quantities of training data. Most current approaches rely on paired low and high-resolution images to train the network in a fully supervised manner. However, such im-

age pairs are not available in real-world applications. To circumvent this fact, the conventional approach has been to downscale images, often with a bicubic kernel, to artificially generate corresponding LR images. This strategy significantly changes the low-level characteristics of the image, by *e.g.* severely reducing the sensor noise. Super-resolution networks trained on downsampled images therefore often struggle to generalize to natural images. The research direction of blind super-resolution [49, 20, 6] does not fully address this setting since it often relies on paired data and constrained image formation models. In this challenge, the aim is instead to learn super-resolution from unpaired data and without any restricting assumptions on the input image formation. This scenario has recently attracted significant interest due to its high relevance in applications [74, 37, 8, 45].

The NTIRE 2020 Challenge on Real-World Image Super-Resolution aims to stimulate research in the direction of real-world super-resolution. No paired reference HR images are available for training. Instead, the participants are only provided the source input images, along with an unpaired set of high-quality images that act as the target quality domain. The challenge consists of two tracks. The source images for Track 1 are generated by performing a degradation operation that is unknown to the participants. This degradation arises from image signal processing methods similar to those found on low-end devices (see Figure 1 for an example). A synthetic degradation allows us to compute reference-based metrics for evaluation. Track 2 employs images taken from a low-quality smartphone camera, with no available ground-truth. For both tracks, the goal is to achieve perceptually pleasing results. The final ranking is therefore performed using a human study.

This challenge is one of the NTIRE 2020 associated challenges on: deblurring [54], nonhomogeneous dehazing [4], perceptual extreme super-resolution [75], video quality mapping [19], real image denoising [1], real-world super-resolution [47], spectral reconstruction from RGB image [5] and demoiréing [73].

2. NTIRE 2020 Challenge

The goals of the NTIRE 2020 Challenge on Real-World Image Super-Resolution is to (i) promote research into weak and unsupervised learning approaches for SR, that jointly enhance the image quality (ii) promote a benchmark protocol and dataset; and (iii) probe the current state-of-the-art in the field. The challenge contains two tracks. Both tracks have the goal of upscaling with factor 4 \times . The competition was organized using the Codalab platform.

2.1. Track 1: Image Processing Artifacts

This track employs the benchmarking strategy described in [45], which employs an artificial degradation operator to

enable reference-based evaluation.

Degradation operator We employ an undisclosed degradation operator which generates structured artifacts commonly produced by the kind of image processing pipelines found on very low-end devices. This type of degradation operator is very different from what has been used in previous challenges [46]. This operation is applied to all source domain images of train, validation and test. According to the rules of the challenge, the participants were not permitted to try to reverse-engineer or with hand-crafted algorithms construct similar-looking degradation artifacts. It was however allowed to try to *learn* the degradation operator using generic techniques (such as deep networks) that can be applied to any other sort of degradations or source of natural images. The reason is that the method as a whole needs to generalize to different types of degradations and input domains.

Data The dataset is constructed following the general strategy used for Track 2 in the previous edition of the challenge [46]. We construct a dataset of source (*i.e.* input) domain training images $\mathcal{X}_{\text{train}} = \{x_i\}$ by applying the degradation operation to the 2650 images of the Flickr2K [67] dataset, without performing any downsampling. The target domain for training $\mathcal{Y}_{\text{train}} = \{y_j\}$ are the original 800 clean high-quality training images from DIV2K. For validation and testing, we employ the corresponding splits from the DIV2K [62] dataset. The source domain images \mathcal{X}_{val} and $\mathcal{X}_{\text{test}}$ are obtained by first downscaling the images followed by the degradation. The Ground Truth images for validation $\mathcal{Y}_{\text{val}}^{\text{gt}}$ and test $\mathcal{Y}_{\text{test}}^{\text{gt}}$ are the original DIV2K images. A visual example for source and target images are provided in Figure 1.

2.2. Track 2: Smartphone Images

Here the task is to super-resolve real-world images obtained from a low-quality smartphone camera. The desired output quality is defined by set of clean high-quality images. We employ the iPhone3 images of the DPED [30] dataset as source domain $\mathcal{X}_{\text{train}}$. For training and validation, we employ the corresponding predefined splits of DPED. As a ground truth to super-resolved images above sensor size does not exist, we use crops of the validation set of DPED for a human perception study. The target domain $\mathcal{Y}_{\text{train}}$ is the same as in Track 1. A visual example for source and target images are provided in Figure 1.

2.3. Challenge phases

The challenge had three phases: (1) Development phase: the participants got training images and the LR images of the validation set. (2) Validation phase: the participants had the opportunity to measure performance using the PSNR and SSIM metrics by submitting their results on the server for Track 1. A validation leaderboard was also available.

(3) Final test phase: the participants got access to the LR test images and had to submit their super-resolved images along with description, code and model weights for their methods.

3. Challenge Results

Before the end of the final test phase, participating teams were required to submit results, code/executables, and fact-sheets for their approaches. From 292 registered participants in Track 1, 19 valid methods were submitted, stemming from 16 different teams. Track 2 had 251 registered participants, of which 15 valid methods were submitted from 14 different teams. Table 1 and 2 report the final results of Track 1 and 2 respectively, on the test data of the challenge. The methods of the teams that entered the final phase are described in Section 4 and the teams’ members and affiliations are shown in Section Appendix A.

3.1. Architectures and Main Ideas

Inspired by the results of the last challenge in AIM 2019 [46] and on the success of recent approaches [45, 18], most top methods pursued a two step approach. The first step aims to learn a network that can transfer *clean* images to the source domain. This network thus learns a degradation operator, adding the kind of noise and corruptions present in the source images. It is then used to generate paired training data for the second step, which involves learning the super resolution network itself. It is generally trained using pairs generated by first downscaling and then applying the learned degradation on images from the target domain set. Many works employed the DSGAN [18] framework from the winner of the AIM 2019 challenge [46] to learn the degradation operator in the first step.

Some of the top methods in this challenge proposed particularly notable alterations and extensions to the general idea described above for learning the degradation network. The AITA-Noah team (Sec. 4.2) employs an iterative approach for Track 1, alternating between learning the degradation and SR network. It also uses an explicit denoising algorithm and train a sharpening network to decrease the blurring effects from the former. Impressionism (Sec. 4.1) is the only team that aims to explicitly estimate the blur kernel in the image, for improved source data generation. For Track 2, it employs the KernelGAN [7] for this purpose. It also aims to explicitly estimate the noise variance using source image patches. This approach led to superior sharpness and quality in the generated SR images for Track 2. There were also some alternative strategies proposed. In particular, the Samsung-SLSI-MSL team (Sec. 4.3) aim to train a robust SR network capable of handling different source domains by randomly sampling a variety of degradations during the training of the SR network.

For the Real-world Super-Resolution setting, the results in the challenge suggest that training strategy and careful degradation modelling is far more important than choice of SR architecture. For the latter, most top methods simply adopted popular architectures, such as the RRDB/ESRGAN [67] and the RCAN [77]. Most methods also included adversarial and perceptual VGG losses, often based on the ESRGAN [67] framework. Brief descriptions of the methods submitted from each team is given in Sec. 4.

3.2. Baselines

We compare methods participating in the challenge with several baseline approaches.

Bicubic Standard bicubic upsampling using MATLAB’s `imresize` function.

RRDB PT The pre-trained RRDB [67], using the network weights provided by the authors. The network was trained with clean images using bicubic down-sampling for supervision. The only objective is the PSNR oriented L1 loss.

ESRGAN Supervised ESRGAN network [67] that is fine-tuned in a fully supervised manner, by applying the synthetic degradation operation used in Track 1. The degradation was unknown for the participants. This method therefore serves as an upper bound in performance, allowing us to analyze the gap between supervised and unsupervised methods. We employ the source $\mathcal{X}_{\text{train}}$ and target $\mathcal{Y}_{\text{train}}$ domain train images respectively. Low-resolution training samples are constructed by first down-sampling the image using the bicubic method and then apply the synthetic degradation. The network is thus trained with real input and output data, which is otherwise inaccessible. As for previous baselines, the network is initialized with the pre-trained weights provided by the authors. Note that no supervised baseline is available for Track 2 since no ground-truth HR images exists.

3.3. Evaluation Metrics

The aim of the challenge is to pursue good image quality as perceived by humans. As communicated to the participants at the start of the challenge, the final ranking was therefore to be decided based on a human perceptual study.

Track 1 For Track 1, the fidelity-based Peak Signal-to-Noise Ratio (PSNR) and the Structural Similarity index (SSIM) [68] was provided on the Codalab platform for quantitative feedback. These metrics are also reported here for the test set. Moreover, we report the LPIPS [76] distance, which is a learned reference-based image quality metric computed as the L^2 distance in a deep feature space. The network itself has been fine-tuned based on image quality annotations, to correlate better with human perceptual opinions. However, this metric needs to be used with great care since many methods employ feature-based losses using ImageNet pre-trained VGG networks, which in its design

| Team | PSNR↑ | SSIM↑ | LPIPS↓ | MOS↓ |
|-------------------|-----------------------|-----------------------|-----------------------|----------------------|
| Impressionism | 24.67 ₍₁₆₎ | 0.683 ₍₁₃₎ | 0.232 ₍₁₎ | 2.195 ₍₁₎ |
| Samsung-SLSI-MSL | 25.59 ₍₁₂₎ | 0.727 ₍₉₎ | 0.252 ₍₂₎ | 2.425 ₍₂₎ |
| BOE-IOT-AIBD | 26.71 ₍₄₎ | 0.761 ₍₄₎ | 0.280 ₍₄₎ | 2.495 ₍₃₎ |
| MSMers | 23.20 ₍₁₈₎ | 0.651 ₍₁₇₎ | 0.272 ₍₃₎ | 2.530 ₍₄₎ |
| KU-ISPL | 26.23 ₍₆₎ | 0.747 ₍₇₎ | 0.327 ₍₈₎ | 2.695 ₍₅₎ |
| InnoPeak-SR | 26.54 ₍₅₎ | 0.746 ₍₈₎ | 0.302 ₍₅₎ | 2.740 ₍₆₎ |
| ITS425 | 27.08 ₍₂₎ | 0.779 ₍₁₎ | 0.325 ₍₆₎ | 2.770 ₍₇₎ |
| MLP-SR | 24.87 ₍₁₅₎ | 0.681 ₍₁₄₎ | 0.325 ₍₇₎ | 2.905 ₍₈₎ |
| Webbzhou | 26.10 ₍₉₎ | 0.764 ₍₃₎ | 0.341 ₍₉₎ | - |
| SR-DL | 25.67 ₍₁₁₎ | 0.718 ₍₁₀₎ | 0.364 ₍₁₀₎ | - |
| TeamAY | 27.09 ₍₁₎ | 0.773 ₍₂₎ | 0.369 ₍₁₁₎ | - |
| BIGFEATURE-CAMERA | 26.18 ₍₇₎ | 0.750 ₍₆₎ | 0.372 ₍₁₂₎ | - |
| BMIPL-UNIST-YH-1 | 26.73 ₍₃₎ | 0.752 ₍₅₎ | 0.379 ₍₁₃₎ | - |
| SVNIT1-A | 21.22 ₍₁₉₎ | 0.576 ₍₁₉₎ | 0.397 ₍₁₄₎ | - |
| KU-ISPL2 | 25.27 ₍₁₄₎ | 0.680 ₍₁₅₎ | 0.460 ₍₁₅₎ | - |
| SuperT | 25.79 ₍₁₀₎ | 0.699 ₍₁₂₎ | 0.469 ₍₁₆₎ | - |
| GDUT-wp | 26.11 ₍₈₎ | 0.706 ₍₁₁₎ | 0.496 ₍₁₇₎ | - |
| SVNIT1-B | 24.21 ₍₁₇₎ | 0.617 ₍₁₈₎ | 0.562 ₍₁₈₎ | - |
| SVNIT2 | 25.39 ₍₁₃₎ | 0.674 ₍₁₆₎ | 0.615 ₍₁₉₎ | - |
| ITA-Noah-A | 24.65 ₍₋₎ | 0.699 ₍₋₎ | 0.222 ₍₋₎ | 2.245 ₍₋₎ |
| ITA-Noah-B | 25.72 ₍₋₎ | 0.737 ₍₋₎ | 0.223 ₍₋₎ | 2.285 ₍₋₎ |
| Bicubic | 25.48 ₍₋₎ | 0.680 ₍₋₎ | 0.612 ₍₋₎ | 3.050 ₍₋₎ |
| ESRGAN Supervised | 24.74 ₍₋₎ | 0.695 ₍₋₎ | 0.207 ₍₋₎ | 2.300 ₍₋₎ |

Table 1. Challenge results for **Track 1**. The top section in the table contains participating methods that are ranked in the challenge. The middle section contains participating approaches that deviated from the challenge rules, whose results are reported for reference but not ranked. The bottom section contains baseline approaches. Participating methods are ranked according to their Mean Opinion Score (MOS).

is very similar to LPIPS. Moreover, some methods directly use the LPIPS distance as a loss for hyper-parameter tuning. We treat LPIPS as an indication of perceptual quality, but not as a metric to decide final rankings.

To obtain a final ranking of the methods, we performed a user study on Amazon Mechanical Turk. For Track 1, where reference images are available, we calculate the Mean Opinion Score (MOS) in the following manner. The test candidates were shown a side-by-side comparison of a sample prediction of a certain method and the corresponding reference ground-truth. They were then asked to evaluate the quality of the SR image w.r.t. the reference image using the 6-level scale defined as: 0 - 'Perfect', 1 - 'Almost Perfect', 2 - 'Slightly Worse', 3 - 'Worse', 4 - 'Much Worse', 5 - 'Terrible'. The images shown to the participants of the study were composed of zoomed crops, as shown in Figure 2. The human study was performed for the top 10 methods according to LPIPS distance, along with 4 baseline approaches.

Track 2 For Track 2, a ground truth reference does not exist due to the nature of the problem. Therefore we used several no-reference based image quality assessment (IQA) metrics. In particular, we report the NIQE [52], BRISQUE [51]

and PIQE [53], using their corresponding MATLAB implementations. Moreover, we report the learned NRQM [48] IQA score. We also report two metrics that summarize the result of the computed IQA metrics. The Perceptual Index PI, previously employed in [32], is calculated as an adjusted mean of NIQE and NRQM. We also compute the mean IQA-Rank by taking the average image-wise rank achieved w.r.t. each of the four IQA metrics. In this case, taking the average rank is preferred over the average value, since the rank is not sensitive to the specific scaling or range of the particular metric.

Since no reference image exists in Track 2, the MOS score as defined for Track 1 cannot be computed. Instead, we compute the Mean Opinion Rank (MOR) by asking the study participants to rank the predictions of several methods in terms of image quality. For each question, the study participants were shown the SR results of all methods in the study for a particular image. These images were then ranked in terms of overall image quality. The MOR is then computed by averaging the assigned rank of each method, over all images and study participants. Since ranking too many entries at once is cumbersome and can lead to inaccurate results, we performed the human study on the top 5 ap-

| Team | NIQE \downarrow | BRISQUE \downarrow | PIQE \downarrow | NRQM \uparrow | PI \downarrow | IQA-Rank \downarrow | MOR \downarrow |
|-------------------|----------------------|----------------------|----------------------|----------------------|----------------------|-----------------------|---------------------|
| Impressionism | 5.00 ₍₁₎ | 24.4 ₍₁₎ | 17.6 ₍₂₎ | 6.50 ₍₁₎ | 4.25 ₍₁₎ | 3.958 | 1.54 ₍₁₎ |
| AITA-Noah-A | 5.63 ₍₄₎ | 33.8 ₍₅₎ | 29.7 ₍₈₎ | 4.23 ₍₈₎ | 5.70 ₍₆₎ | 7.720 | 3.04 ₍₂₎ |
| ITS425 | 8.95 ₍₁₈₎ | 52.5 ₍₁₈₎ | 88.6 ₍₁₈₎ | 3.08 ₍₁₈₎ | 7.94 ₍₁₈₎ | 14.984 | 3.30 ₍₃₎ |
| AITA-Noah-B | 8.18 ₍₁₇₎ | 50.1 ₍₁₂₎ | 88.0 ₍₁₇₎ | 3.23 ₍₁₅₎ | 7.47 ₍₁₇₎ | 13.386 | 3.57 ₍₄₎ |
| Webbzhou | 7.88 ₍₁₅₎ | 51.1 ₍₁₅₎ | 87.8 ₍₁₆₎ | 3.27 ₍₁₄₎ | 7.30 ₍₁₅₎ | 12.612 | 4.44 ₍₅₎ |
| Relbmag-Eht | 5.58 ₍₃₎ | 33.1 ₍₃₎ | 12.5 ₍₁₎ | 6.22 ₍₂₎ | 4.68 ₍₂₎ | 4.060 | - |
| MSMers | 5.43 ₍₂₎ | 38.2 ₍₇₎ | 20.5 ₍₃₎ | 5.22 ₍₅₎ | 5.10 ₍₃₎ | 5.420 | - |
| MLP-SR | 6.45 ₍₈₎ | 30.6 ₍₂₎ | 29.0 ₍₆₎ | 6.12 ₍₃₎ | 5.17 ₍₄₎ | 5.926 | - |
| SR-DL | 6.11 ₍₅₎ | 33.5 ₍₄₎ | 29.4 ₍₇₎ | 5.24 ₍₄₎ | 5.43 ₍₅₎ | 6.272 | - |
| InnoPeak-SR | 7.42 ₍₁₃₎ | 39.3 ₍₈₎ | 21.5 ₍₄₎ | 5.12 ₍₆₎ | 6.15 ₍₉₎ | 7.716 | - |
| QCAM | 6.21 ₍₆₎ | 44.2 ₍₉₎ | 49.6 ₍₉₎ | 4.10 ₍₁₀₎ | 6.05 ₍₈₎ | 8.304 | - |
| SuperT | 6.94 ₍₁₀₎ | 50.2 ₍₁₃₎ | 75.1 ₍₁₁₎ | 4.23 ₍₉₎ | 6.35 ₍₁₀₎ | 9.612 | - |
| KU-ISPL | 6.79 ₍₉₎ | 45.1 ₍₁₀₎ | 61.6 ₍₁₀₎ | 3.60 ₍₁₃₎ | 6.59 ₍₁₂₎ | 10.152 | - |
| BMIPL-UNIST-YH-1 | 7.03 ₍₁₂₎ | 50.2 ₍₁₄₎ | 81.5 ₍₁₃₎ | 3.70 ₍₁₂₎ | 6.66 ₍₁₃₎ | 12.218 | - |
| BIGFEATURE-CAMERA | 7.45 ₍₁₄₎ | 49.2 ₍₁₁₎ | 87.1 ₍₁₄₎ | 3.23 ₍₁₆₎ | 7.11 ₍₁₄₎ | 13.784 | - |
| Samsung-SLSI-MSL | 6.25 ₍₇₎ | 37.3 ₍₆₎ | 26.0 ₍₅₎ | 4.31 ₍₇₎ | 5.97 ₍₇₎ | 6.662 | - |
| Bicubic | 7.97 ₍₁₆₎ | 52.0 ₍₁₇₎ | 87.2 ₍₁₅₎ | 3.16 ₍₁₇₎ | 7.40 ₍₁₆₎ | 14.532 | 6.04 ₍₆₎ |
| RRDB | 7.01 ₍₁₁₎ | 51.3 ₍₁₆₎ | 76.0 ₍₁₂₎ | 4.06 ₍₁₁₎ | 6.48 ₍₁₁₎ | 10.042 | 6.06 ₍₇₎ |

Table 2. Challenge results for **Track 2**. The top section in the table contains participating methods that are ranked in the challenge. The middle section contains participating approaches that deviated from the challenge rules, whose results are reported for reference but not ranked. The bottom section contains baseline approaches. Participating methods are ranked according to their Mean Opinion Rank (MOR).

proaches along with two baselines. As we did not find any of the IQA metrics previously discussed to correlate well with perceived image quality, the initial selection of top 5 methods was performed using a purely visual comparison performed by the challenge organizers. The top 5 methods were selected by assessing sharpness, noise, artifacts, and overall quality. The MOR scores were then computed using Amazon Mechanical Turk.

3.4. Track 1: Image Processing Artifacts

Here we present the results for Track 1. All experiments presented were conducted on the test set. The results are shown in Table 1. The Impressionism team achieves the best result, with a 9.5% better MOS than the second entry, namely Samsung-SLSI-MSL. Both these teams take a more direct approach for simulating degradations for supervised SR learning. While Samsung-SLSI-MSL sample random noise distributions and down-scaling kernels, Impressionism aim to estimate the kernel and noise statistics. The following three approaches: BOE-IOT-AIBD, MSMers, and KU-ISPL, employ CycleGAN [12] or DSGAN [18] based methods to learn the degradation operator. Also the AITA-Noah team follows this general strategy, achieving impressive MOS results. However, their methods are not ranked in Track 1 since source domain images from the test set was used for training, which is against the rules of the challenge. Notable are also the results of ITS425, who achieve the second best PSNR and best SSIM, while preserving good per-

ceptual quality. Also the third-ranked method BOE-IOT-AIBD achieves very impressive PSNR and SSIM.

When comparing with the previous edition of the challenge [46], the performance of the proposed method has improved substantially. In [46], most method achieved similar or worse results than simple Bicubic interpolation. Here, all top-10 approaches achieved better MOS than the Bicubic baseline. Moreover, while a large gap to supervised methods was reported in [46], in this year challenge, the winning Impressionism method even beats the ESRGAN baseline, which is trained with full supervision. While this can also be partly explained by other modifications and hyper-parameter settings, it clearly demonstrates that the performance gap to supervised SR methods is significantly narrower. Visual results for all methods in Figure 2.

3.5. Track 2: Smartphone Images

Quantitative results for Track 2 are reported in Table 2. In this track, the Impressionism method outperforms other approaches by a large margin in the human study (MOR). This is also confirmed in the visual examples shown in Figure 3. The generated images are superior in sharpness compared to those of other approaches. Moreover, the SR images contain almost no noise and few artifacts. While AITA-Noah and ITS425 also generate clean images, they lack the sharpness and detail of Impressionism. We believe this to be largely due to the kernel estimation performed in the latter approach, employing KernelGAN for this purpose.

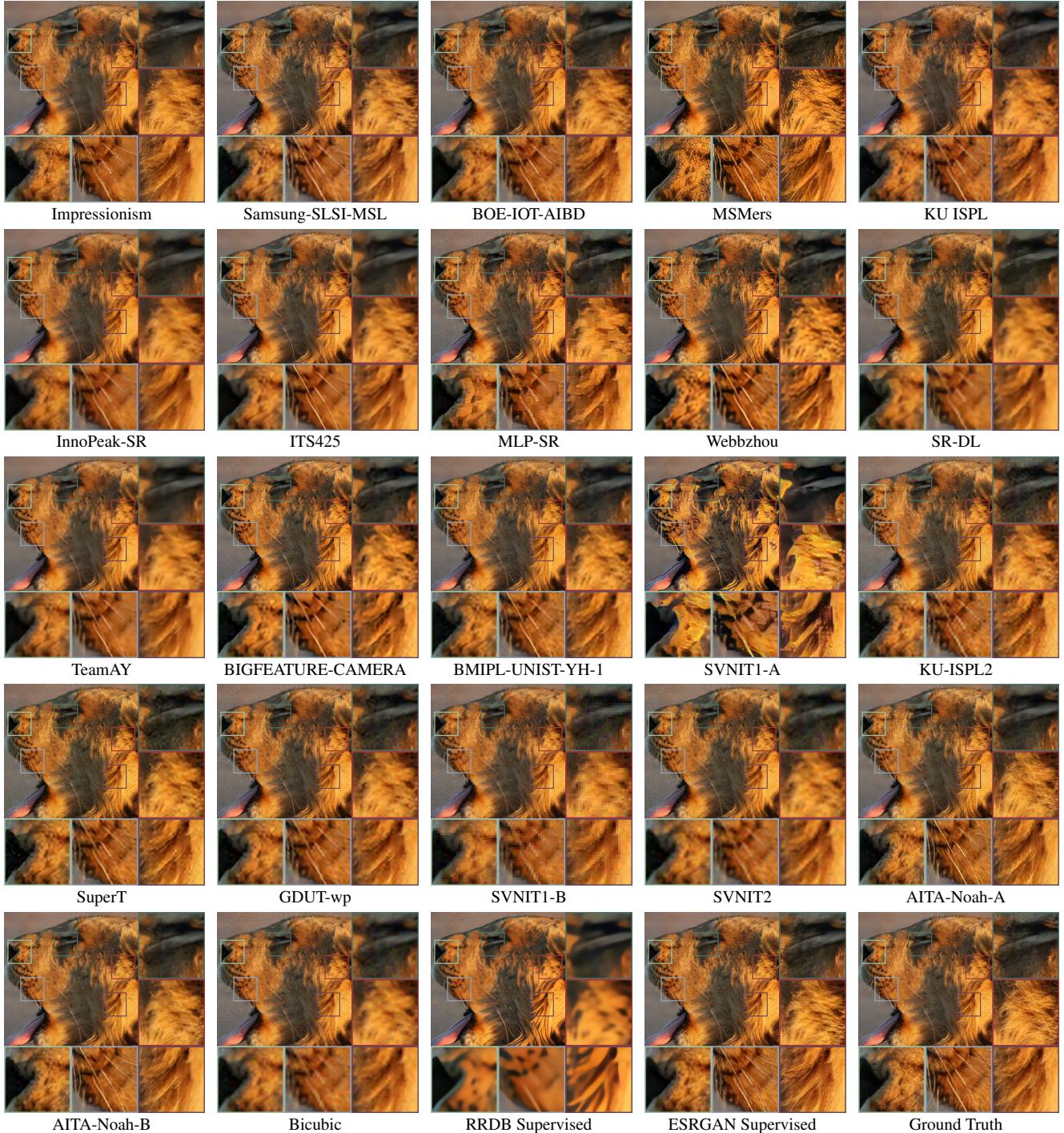


Figure 2. Qualitative comparison between the participating approaches for Track 1. (4× super-resolution)

This allows the SR network to take the pointspread function of the specific camera sensor into account.

We observe that Impressionism also achieves the best average IQA-Rank. However, note that while the Relbmag-Eht team achieves a similar IQA-Rank, their result severely suffers from a structured noise pattern. This suggests that standard IQA metrics are not well suited as evaluation cri-

teria in this setting and data. Interestingly, the Samsung-SLSI-MSL team employed the paired DSLR images provided by [30]. This approach is therefore not ranked in this track. However, this approach does still not achieve close to the same level of sharpness as Impressionism.

Despite being the first challenge of its kind, the top participating teams achieved very impressive results in this dif-

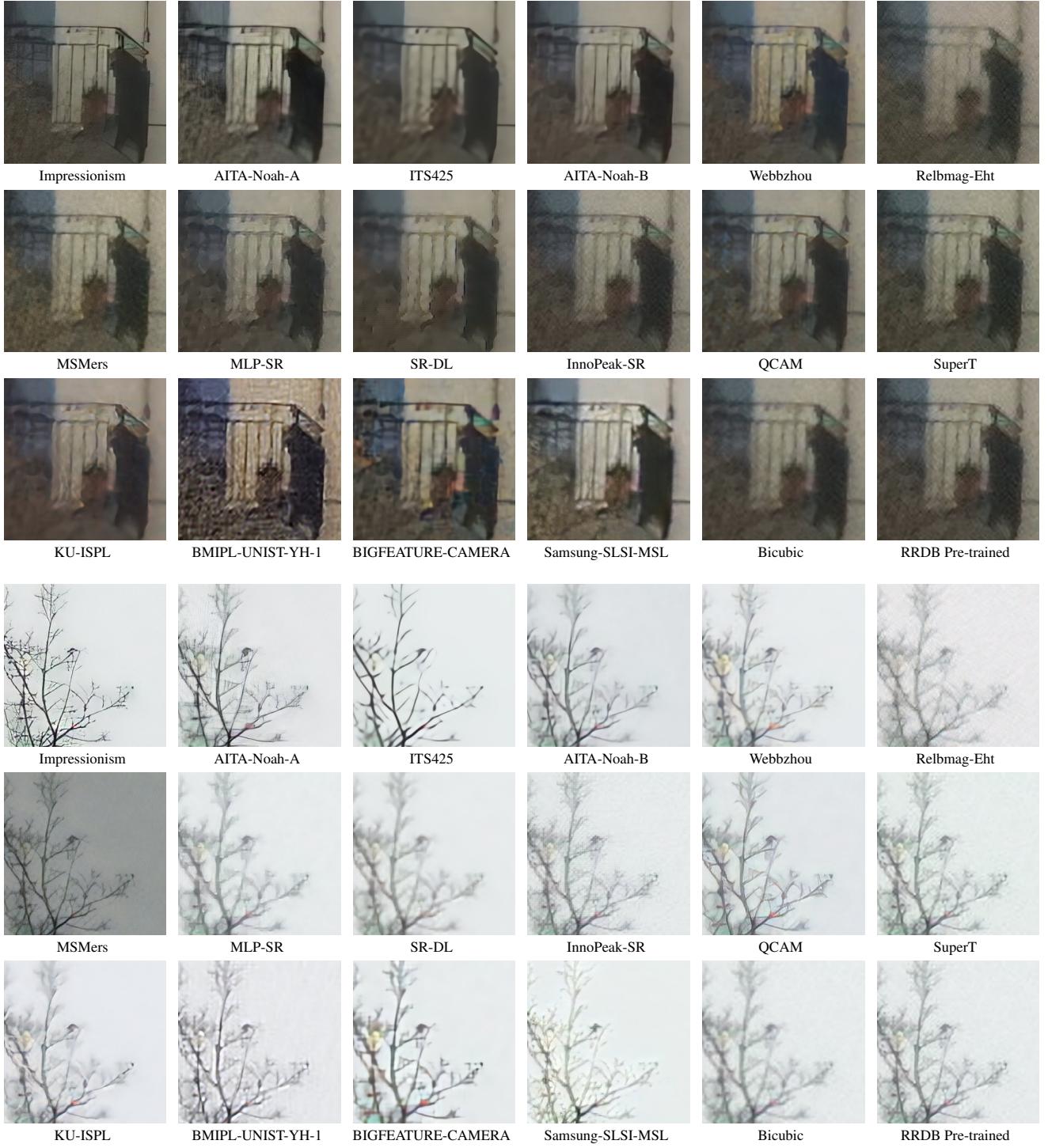


Figure 3. Qualitative comparison between the participating approaches for Track 2. (4× super-resolution)

ficult real-world setting, where no reference data is available. In particular, the Impressionism team achieves not only a higher resolution image, but also substantially better image quality than the source image taken by the camera.

4. Challenge Methods and Teams

This sections give brief descriptions of the participating methods. A summary of all participants is given in table 4.1.

4.1. Impressionism

The team Impressionism proposes a novel framework, introduced in [35], to improve the robustness of the super-resolution model on real images, which usually fails when trained on bicubic downsampled data. To generate more realistic LR images, they design a real-world degradation process that maintains important original attributes. Specifically, they focus on two aspects: 1) The blurry LR image is obtained by downsampling High-Resolution (HR) images with estimated kernels from real blurry images. 2) The real noise distribution is restored by injecting collected noise patches from real noisy images. From the real-world (source domain) dataset \mathcal{X} and the clean HR (target domain) dataset \mathcal{Y} , the team thus aims to construct domain-consistent data $\{\mathbf{I}_{LR}, \mathbf{I}_{HR}\} \in \{\mathcal{X}, \mathcal{Y}\}$.

Clean-up Since bicubic downsampling can remove high-frequency noise, they directly do a bicubic downsampling on the image from \mathcal{X} to obtain more HR images. Let $\mathbf{I}_{src} \in \mathcal{X}$, and \mathbf{k}_{bic} be the ideal bicubic kernel. Then the image is downsampled with a clean-up scale factor s as $\mathbf{I}_{HR} = (\mathbf{I}_{src} * \mathbf{k}_{bic}) \downarrow_s$. Then the images after downsampling are regarded as clean HR images, that is $\mathbf{I}_{HR} \in \mathcal{Y}$.

Downsampling The team performs downsampling on the clean HR images using the estimated kernels by KernelGAN [7]. The downsampling process is a cross-correlation operation followed by sampling with stride s ,

$$\mathbf{I}_D = (\mathbf{I}_{HR} * \mathbf{k}_i) \downarrow_s, \quad i \in \{1, 2, \dots, m\}, \quad (1)$$

where \mathbf{I}_D denotes the downsampled image, and \mathbf{k}_i refers to the specific blur kernel.

Noise Injection Mere estimation of the blurry kernel cannot accurately model the degradation process of \mathcal{X} . By observing the real data, they find that the noise is usually combined with content of the image. In order to decouple noise and content, they design a filtering rule to collect noise patches $\{\mathbf{n}_i, i \in \{1, 2 \dots l\}\}$ with their variance in a certain range $\sigma^2(\mathbf{n}_i) < v$, where $\sigma^2(\cdot)$ denotes the variance, and v is the maximum value of variance. Then these patches will be added to \mathbf{I}_D as,

$$\mathbf{I}_{LR} = \mathbf{I}_D + \mathbf{n}_i, \quad i \in \{1, 2 \dots l\}. \quad (2)$$

After downsampling HR images with the estimated kernels and injecting collected noise, they obtain $\mathbf{I}_{LR} \in \mathcal{X}$.

Network Details Based on ESRGAN [67], they train a super-resolution model on constructed paired data $\{\mathbf{I}_{LR}, \mathbf{I}_{HR}\} \in \{\mathcal{X}, \mathcal{Y}\}$. Three losses are applied to training including pixel loss L_1 , perceptual loss L_{per} , and adversarial loss L_{adv} . Different from default setting, they use patch

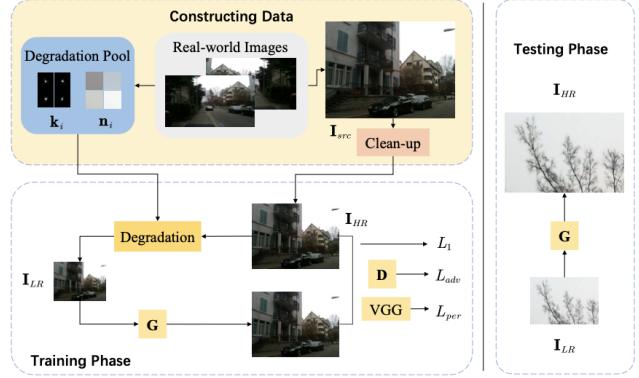


Figure 4. Overview of the method by the **Impressionism** team.

discriminator [34] instead. Overall, the final training loss is as follows:

$$L_{total} = \lambda_1 * L_1 + \lambda_{per} * L_{per} + \lambda_{adv} * L_{adv}, \quad (3)$$

where λ_1 , λ_{per} , and λ_{adv} are set as 0.01, 1, and 0.005 empirically.

4.2. AITA-Noah

This method, which is detailed in [10], adopts the idea of learning the degradation operator in order to synthetically generate paired training data for SR network. For Track 1, an approach termed *Iterative Domain Adaptation* is developed. The source training data \mathcal{X}_{tr} and downsampled target training data $\mathcal{Y}_{tr\downarrow}$ are first processed with a denoising algorithm (Non-local Means), denoted D . The sets $D(\mathcal{Y}_{tr\downarrow})$ and $\mathcal{Y}_{tr\downarrow}$ are then used to train a sharpening network S , in a fully supervised manner. When applied to the source data, $S(D(x_{tr}))$ generates images that are clean and sharp. This set can then be used to train a degradation operator G , using pairs from $S(D(\mathcal{X}_{tr}))$ and \mathcal{X}_{tr} . This is then used to train a super-resolution network SR using pairs generated by $G(\mathcal{Y}_{tr\downarrow})$ and \mathcal{Y}_{tr} . The approach then proceeds by iteratively improving the degradation model G using pairs $f(\mathcal{X}_{tr})$ generated by the current SR model f and \mathcal{X}_{tr} , and improving the super-resolution model f using pairs $G(\mathcal{Y}_{tr})$ generated by the current degradation operator G and \mathcal{Y}_{tr} . In practice, the team used the 100 source validation images and 100 source test images as \mathcal{X}_{tr} . The team is not ranked in track 1, since according to the challenge rules, the test data should not be used during training, even in unpaired form.

For Track 2, the team adopts the CycleSR framework [24, 45] to generate degrade images. As illustrated in Fig. 5, this framework is composed of two stages: 1) unsupervised image translation between real LR images \mathcal{X}_{tr} and synthetic LR images, *i.e.*, 4× bicubic downsampled HR images \mathcal{Y}_{tr} , denoted by $\mathcal{Y}_{tr\downarrow}$; 2) supervised super-resolution from degraded LR images $\hat{\mathcal{Y}}_{tr\downarrow}$ to get $\hat{\mathcal{Y}}_{tr}$. In detail, the ap-

| Team Name | Username in Codalab | Additional Data | Track 1 | | Track 2 | |
|-------------------|---------------------|---|---------------|---------------|---------------|---------------|
| | | | Traintime [h] | Runtime [sec] | Traintime [h] | Runtime [sec] |
| AITA-Noah-A | AITA | Track 1 and 2: AIM-2019 pretrained model. Track 2: only use external 400 div8k images. | 8 | 0.5 | 8 | 0.5 |
| AITA-Noah-B | Noah.TerminalVision | AIM-2019 pretrained ESRGAN-FS model. | 8 | 5 | 8 | 0.3 |
| BIGFEATURE_CAMERA | conson0214 | DSGAN for LR-HR pairs, DF2K to pre-train SR model. | 22 | 0.25 | 22 | 0.25 |
| BMIPL_UNIST_YH_1 | syh | RCAN Super Resolution model | 32 | 40 | 12 | 40 |
| BOE-IOT-AIBD | eastworld | 739 pixels.com images, downsized to 2K | 264 | 38.20 | no | no |
| GDUT-wp | HouseLee | - | 10 | 0.85 | no | no |
| ITS425 | Ziyao_Zong | - | 24 | 1.34 | 24 | 1.24 |
| Impressionism | xiaozhongji | RRDB_PSNR_x4.pth released by the ESRGAN authors | 12 | 1.3 | 32 | 0.9 |
| InnoPeak_SR | qizhangTiTi | 10,000 collected images | 12 | 0.15 | 12 | 0.15 |
| KU-ISPL2 | Kanghyu_Lee | VGG19 was used for VGG loss | 2 | 0.02 | no | no |
| KU_ISPL | gkим | VGG-19 model for perceptual loss | 168 | 6.48 | 168 | 4.11 |
| MLP_SR | raoumer | - | 28.57 | 1.289 | 0 | 967 |
| MSMers | huayan | CycleGan, RCAN | 72 | 0.483 | 63 | 0.343 |
| QCAM | tkhu | AIM2019 | no | no | 15 | 0.21 |
| Relbmag_Eht | Timothy_Cilered | - | no | no | 8.9 | 1.09 |
| SR_DL | ZhiSong_Liu | - | 15 | 4 | 15 | 1 |
| SVNIT1-A | kalpesh_svnit | - | 50 | 1.09 | no | no |
| SVNIT1-B | Kishor | - | 50 | 0.85 | no | no |
| SVNIT2 | vishalchudasama | - | 50 | 0.92 | no | no |
| Samsung_SLSI_MSL | Samsung_SLSI_MSL | Flickr2K for Track 1, DPED for Track 2. | 72 | 1 | 24 | 1 |
| SuperT | tongtong | DIV2K | 48 | 0.64 | 48 | 0.64 |
| TeamAY | nmhkahn | - | 100 | 20 | no | no |
| Webbzhou | Webbzhou | - | 60 | 0.5 | 60 | 0.5 |

Table 3. Information about the participating teams in the challenge.

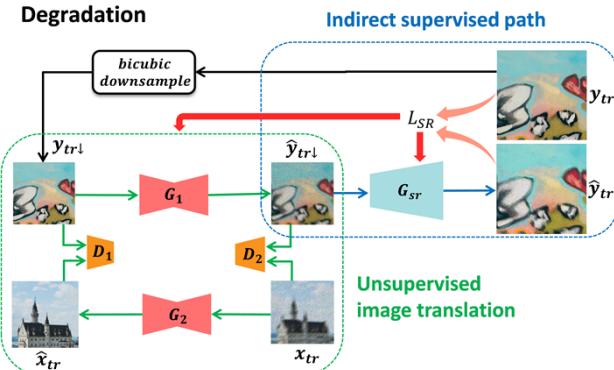


Figure 5. Overview of the CycleSR method used by AITA-Noah to learn the degradation operation for Track 2.

proach first takes the unsupervised image translation model CycleGAN [81] for mapping between domain \mathcal{X}_{tr} and $\mathcal{Y}_{tr}\downarrow$. An SR module SRResNet is employed after CycleGAN to super-resolve $\hat{\mathcal{Y}}_{tr}\downarrow$ to get $\hat{\mathcal{Y}}_{tr}$ and compute the loss L_{SR} with ground truth \mathcal{Y}_{tr} . Hence, with an image translation model and an SR module together and a joint training strategy, we are able to train a model that super-resolves real LR images to HR images with an indirect supervised path. Compared with degradation directly using original CycleGAN, benefiting from the pixel-wise feedback of the SR module, CycleSR can alleviate color and brightness changes during degradation.

In both tracks, the same super-resolution architecture, based on the ESRGAN is used. The team furthermore use an LR-conditional frequency-separation discriminator to train the model and employ AutoML to tune the loss

weights, employing LPIPS [76] and NIQE [52] as objective. Two versions of this approach was submitted, with the significant differences as follows:

AITA-Noah-A For Track 1, this version uses the method described above. For Track 2, it includes an extra 400 images selected from DIV8K [22] in the target domain set \mathcal{Y}_{tr} to improve data diversity.

AITA-Noah-B For Track 1, this approach additionally uses an ensemble fusion strategy (*i.e.* running inference on the vertical flipped/horizontal flipped/transposed images of the original input, and then average the results), in addition to above. For Track 2, no extra data was used and no adversarial loss was used during training the ESRGAN model (*i.e.* only RRDBNet was used).

4.3. Samsung-SLSI-MSL

For Track 1, this team aims to train a generic SR model that is robust to various image degradations, which can therefore be applied in real-world scenarios without knowledge of the specific degradation operator. This is performed by sampling diverse degradation types during training. The strategy proposed in the blind denoising method [59] is extended by adding downscaling and blur. The training set is generated by sampling different downscaling (*e.g.* bilinear, nearest neighbor or bicubic), blur kernels (Gaussian kernel with different sigma), and noise distributions (additive Gaussian, Poisson, Poisson-Gaussian with randomly sampled parameters). The SR model consists of the RCAN [77] architecture, which is trained with a GAN loss while emphasizing the perceptual losses. To further improve the perceptual quality, they deploy an ensemble of two different GANs, and use cues from the image luminance and adjust to

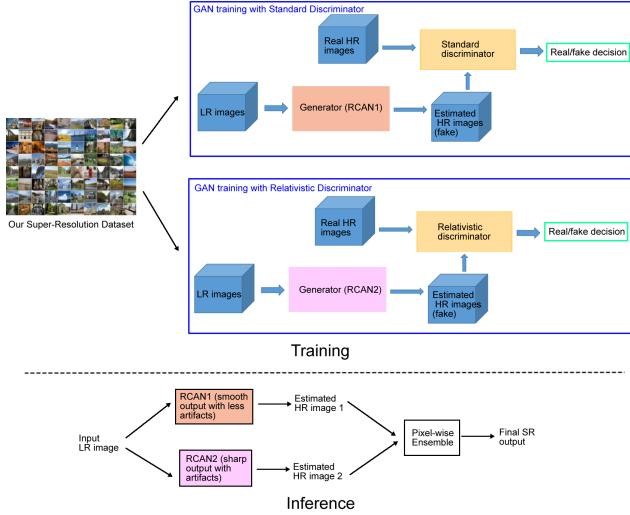


Figure 6. Overview of the SR method used by **Samsung-SLSI-MSL**.

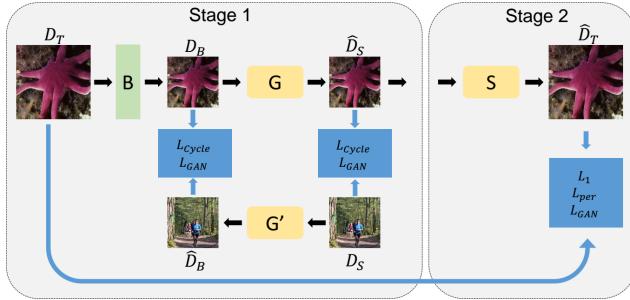


Figure 7. Overview of the method by the **MSMers** team.

generate better HR images at low-illumination. The workflow is given in Fig. 6.

For Track 2, real world SR on images captured by mobile devices, the same GANs are trained by weak supervision on a mobile SR training set that they constructed to have LR-HR image pairs, from the DPED dataset which provides registered mobile-DSLR images at the same scale [30]. They use the mobile images as LR, and apply the track 1 generic SR model on the paired DSLR images to create super resolved HR images with good perceptual quality. This method is considered as a kind of *Supervised* approach, and does not compete with the other participants in Track 2. Details about the proposed method can be found in [25].

4.4. MSMers

This method takes inspiration from [45], developing a two-stage approach. First, a degradation operator is learned in an unsupervised manner. This is then used to generate paired data for the second stage, in which the SR network is learned. Specifically, CycleGAN [81] is adopted in the first stage to learn a mapping from bicubic downsampled HR to real LR. To keep the color consistent, the weight of the iden-

tity loss is increased in the setting. As for the second stage, RCAN [77] is used to super-resolve the LR image, which is first trained on L1 loss. On top of that, perceptual loss and adversarial loss are added for better perceptual quality. Specifically, we use features of VGG19 relu5-1 layer to compute a perceptual loss and the WGAN-GP [23] as adversarial loss. The method is visualized in Figure 7.

4.5. BOE-IOT-AIBD

This team aims to learn the degradation operator in order to generate paired SR training samples. To this end, it employs solution provided by DSGAN [18] to artificially generate LR images, as shown in Figure 8a. These are then used to train an SR model. For this, it uses the modified MGBv2 [50] network, proposed in the winning solution of the AIM ExtremeSR challenge [21]. It is adapted to $4\times$ upscaling by using a triple-V cycle (instead of the W-cycle) and adding multi-scale denoising modules as shown in Figure 8b. During inference, an overlapping patch approach is used to further allow upscaling of large images. The training strategy employs a multiscale loss, combining distortion and perception losses on the output images. Model selection was performed by selecting low NIQE results on validation set and human tests based on ITU-T P.910. An additional set of 739 collected images for training. The team only participated in Track 1.

4.6. InnoPeak-SR

This approach does not directly address the unavailability of paired training data. Instead, it aims to develop a robust architecture capable of generalizing to the degra-

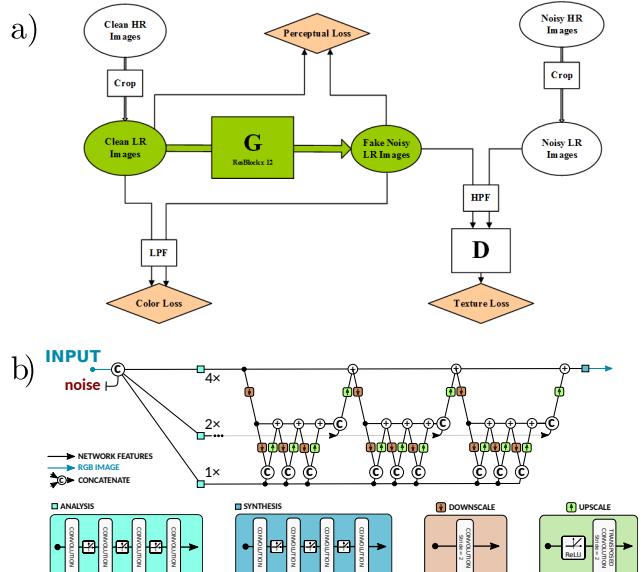


Figure 8. Overview of the DSGAN [18] based method used by the **BOE-IOT-AIBD** team to learn the degradation operation.

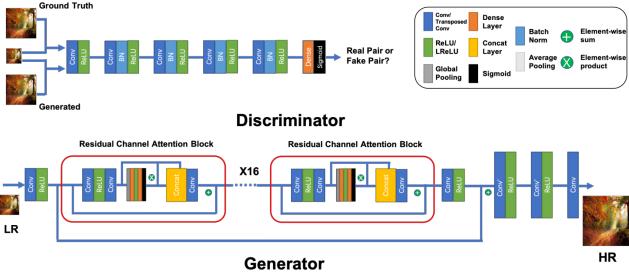


Figure 9. Architectures employed by the **InnoPeak-SR** team.

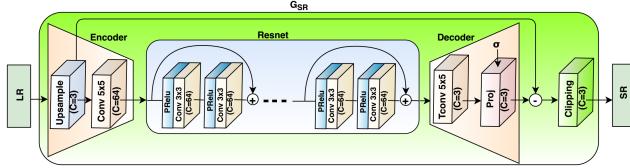


Figure 10. SR architecture employed by the **MLP-SR** team.

tions present in the real-world setting, while trained using standard strategies. The SR network consists of a residual channel attention generator, visualized in Figure 9. It mainly consists of four parts: shallow feature extraction, residual channel attention feature extraction, upscale module, and reconstruction. The discriminator network is implemented using four repeated 4×4 convolution layers, followed by BatchNorm and ReLU. The networks are trained in a standard GAN fashion. The generator additionally uses L_1 , VGG, SSIM and gradient losses. The authors additionally used 10000 images from the ImageNet dataset for training.

Details about the proposed method can be found in [9]

4.7. ITS425

This team focus on improving the SR network architecture. The image degradation operator is first learned using an improved version of the DSGAN [18], by using a smaller generator model than that of the original work. Unlike other methods, this team also aims to improve the quality of the target domain images. This is performed by training denoising and detail enhancement models to improve the target domain HR training images. The SR models is based on the RDN architecture [79]. It is modified by using the add operation instead of concatenate, which not only reduces the amount of calculations of the model but also reduces the high-level information that is passed back to the final layers.

4.8. MLP-SR

This team follow a two stage approach. First, a DSGAN [18] (winner of the AIM2019 RWSR challenge) network and training strategy is employed to learn the image degradation mapping. This is then used to generate paired SR training data for the second stage. The team proposes a

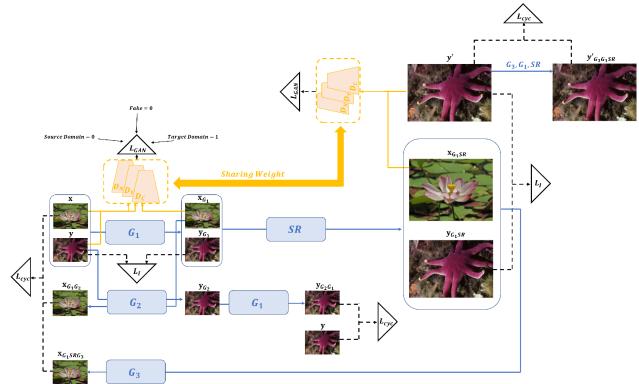


Figure 11. Overview of the learning approach proposed by the **KU-ISPL** team.

SR architecture, shown in Figure 10, inspired by a physical image formation model. It uses a encoder-decoder structure. The inner ResNet consists of 5 residual blocks with two pre-activation Conv layers. The pre-activation is the parametrized rectified linear unit (PReLU). The trainable projection layer [42] inside Decoder computes the proximal map with the estimated noise standard deviation and handles the data fidelity and prior terms. The noise realization is estimated in the intermediate ResNet that is sandwiched between Encoder and Decoder. The estimated residual image after Decoder is subtracted from the LR input image. Reflection padding is also used before all Conv layers to ensure slowly-varying changes at the boundaries of the input images. The generator structure can also be described as the generalization of one stage TNRD [11] and UDNet [42] that have good reconstruction performance for image denoising problem. For the discriminator, it employs the architecture used in SRGAN [41], with the relativistic loss used in ESRGAN [67]. In addition, L_1 , Total-Variation and VGG losses are used.

4.9. KU-ISPL

This team propose an un-paired GAN-based framework [36]. It consists of three generators, one SR model and three discriminators. The overall architecture is visualized in Figure 11. The generators G_1 , G_2 , and G_3 constitute a modified CinCGAN [74]. Residual networks are used for these architectures. G_3 further downsamples the image by a factor of 4. The SR model is based on ESRGAN [67]. Bilinear upsampling is introduced into the architecture to preserve details and avoid checkerboard patterns induced by the transposed convolution module. The three discriminators D_N , D_C , and D_Y are trained with different losses: adversarial noise loss, adversarial color loss, and adversarial texture loss respectively. The D_N uses a raw image, which contains noise signal. The D_C and D_Y employ a Gaussian blurred image and a grayscale image, respectively, as

in WESPE [31]. To improve performance of the discriminators, source domain images are used when the discriminators are trained. Instead of classifying real or fake, the discriminator distinguishes between source and target domain images. The generator is trained to make target domain-like fake images and the discriminator is trained to classify fake images as a source domain image. The cycle consistency and identity loss each consist of three losses: a pixel-wise L_1 loss, a VGG perceptual loss, and an SSIM loss.

4.10. Webbzhou

This team aims to first learn the degradation process in order to generate data for a second-stage SR network training. The degradation learning is based on the frequency separation in DSGAN [18]. Furthermore, in order to alleviate the color shift in degradation process, the team proposed a generator based on Color Attention Residual Block (CARB) [80]. In addition, the team modified the discriminator of ESRGAN [67] which treats high frequency and low frequency separately. Finally, an EdgeLoss with Canny operator is constructed to further enhance details of edge.

4.11. SR-DL

The team propose a joint image denoising and super-resolution model by using generative Variational AutoEncoder (dSRVAE) [44]. It includes two parts: a Denoising AutoEncoder (DAE) and a Super-Resolution Sub-Network (SRSN). With the absence of target images, a simple discriminator is trained together with the autoencoder to encourage the SR images to pick up the desired visual pattern from the reference images. During the training, Denoising AutoEncoder (DAE) is trained first by using source image training set. Then the Super-Resolution Sub-Network (SRSN) is attached as a small head to the DAE which forms the proposed dSRVAE to output super-resolved images. Together with dSRVAE, a simple convolutional neural network is used as a discriminator to distinguish whether generated SR images are close to the original input images.

The method is visualized in Figure 12. The proposed dSRVAE network first uses the encoder to learn the latent vector of the clean image. A Gaussian model randomly samples from the latent vector to the decoder. The input noisy LR image is also included as a conditional constraint to supervise the reconstruction of the decoder. Combining both noisy image features and latent features, the decoder learn the noise pattern. Finally, the estimated clean image is obtained by subtracting the estimated noise from the input noisy image. At the second stage, Super-Resolution Sub-Network (SRSN) is added to the end of the Denoising AutoEncoder to take both bicubic interpolated original clean and estimated denoised images as input to generation superresolution result. Since there is no ground truth of super-resolved images, a discriminator is trained to distinguish the

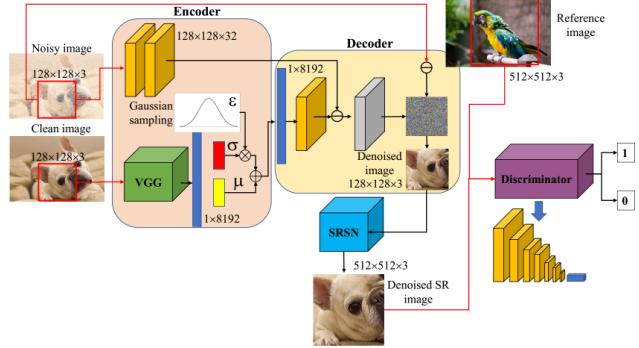


Figure 12. Overview of the method proposed by the **SR-DL** team.

super-resolution results and cropped reference image. The balance is achieved when the discriminator cannot distinguish between reference and denoised SR image.

4.12. TeamAY

This team proposes a simple but strong method for unsupervised SR (SimUSR). Their approach is based on the zero-shot super-resolution (ZSSR) [60] which trains the image-specific network at runtime using only a single given test image \mathbf{I}_{LR} . The ZSSR enables to optimize the model even if high-resolution images are not accessible. However, ZSSR suffers from high runtime latency and inferior performance compared to the supervised SR methods. To mitigate such issues, this team first slightly relax the constraint of ZSSR and assumes that it is relatively easy to collect the LR images, $\{\mathbf{I}_{LR_1}, \dots, \mathbf{I}_{LR_N}\}$. Thanks to this assumption, they can convert fully unsupervised SR into the supervised learning regime by generating multiple pseudo-pairs $\{(\mathbf{I}'_{LR_1}, \mathbf{I}'_{HR_1}), \dots, (\mathbf{I}'_{LR_N}, \mathbf{I}'_{HR_N})\}$ by

$$(\mathbf{I}'_{LR_k}, \mathbf{I}'_{HR_k}) = (\mathbf{I}^{son}_{LR_k}, \mathbf{I}^{father}_{LR_k}), \text{ for } k = 1 \dots N.$$

where $\mathbf{I}^{son}_{LR} = \mathbf{I}_{LR} \downarrow_{s,k}$ and $\mathbf{I}^{father}_{LR} = \mathbf{I}_{LR}$.

Though this is a very simple correction, their modification brings several benefits: It allows their framework to exploit every benefit of supervised learning. For instance, unlike ZSSR, their SimUSR can utilize recently developed network architectures and training techniques that provide huge performance gains. In addition, since the online (runtime) training is not necessary, SimUSR can significantly reduce its runtime latency. For the NTIRE 2020 challenge, they use pretrained RCAN [77] (on bicubic $\times 4$ scale) as a backbone model of SimUSR. Also, they attach ad-hoc denoiser (BM3D [27]) before train the SimUSR method. Details about the proposed method can be found in [55].

4.13. Bigfeature-Camera

This method use DSGAN [18] to learn the degradation, used for generating paired training data. In the second stage

a RNAN [78] based SR network is trained. It is modified to handle multiple scales and by adding a contrast channel attention layers [77] along with local attention blocks.

4.14. BMIPL-UNIST-YH

This method focus on how to train on unpaired data. Similar to [45], a CycleGAN is used to learn the degradation. In the second stage, and RCAN [77] SR architecture is trained on generated data.

4.15. SVNIT1

This team combines self- and unsupervised strategies to train the SR network without supervision. For the self-supervised part, the LR input is upsampled bicubically and used for a pixel-wise loss. The unsupervised losses consist of a Total-Variation loss and a deep image quality loss. For the latter loss, a pre-trained quality assessment network was used. Details about the proposed method can be found in [57]. Two versions of this approach was submitted:

SVNIT1-A In addition to above, this version employs an adversarial loss on the SR output. The discriminator architecture is inspired by [58].

SVNIT1-B Instead of a descriminator, this variant Variational Encoder which follows the architectural guidelines in [58].

4.16. SVNIT2

This method uses cyclic consistency between an SR network and a downscaling network. Two generator are trained: the SR generator going from LR to HR and the downscaling generator going from HR to LR. In addition to cycle consistency, the VGG loss, GAN loss, and a learned image quality loss is employed.

4.17. KU-ISPL2

This team base their approach on SRGAN [41]. This is extended with a multi-scale convolutional block, that combines the results of convolutions with different kernel sizes.

4.18. SuperT

This method uses a balanced Laplacian pyramid network [39] for progressive image super-resolution. For training, both degraded and clean images are used with standard downsampling them for training data generation.

4.19. GDUT-wp

This method uses an ensemble of SRResNets trained on bicubic downsampled data. The idea is that by selecting the best from an ensemble, the effect of random artifacts can be reduced.

4.20. MLP-SR

This method is based on the DSGAN [18] approach. The loss of the super-resolution method consists of a VGG, GAN, TV and L1 loss. To improve the fidelity, they further used a ensemble method at test time [64]. Details about the proposed method can be found in [66].

4.21. Relbmag-Eht

Instead of generating ‘fake’ natural image as DSGAN [18], this team aims to improve this method to aggregate this paring procedure into the super-resolution model. To supervise this matching from HR or bicubic images to natural images, a module with discriminators both in the LR and HR phase is proposed. It allows the downsampling model to learn from upsampling results. The ESRGAN [67] is used as SR model.

4.22. QCAM

This work fine-tunes a pretrained SR model on real data using only supervision in the low-resolution. That is, it aims to minimize the loss $\min_{\theta} \|D(f_{\theta}(x)) - x\|^2$ for source images x . Here, f_{θ} is the SR model with parameters θ and D is the bicubic downsampling operation.

5. Conclusions

This paper presents the setup and results of the NTIRE 2020 challenge on real world super-resolution. Contrary to conventional super-resolution, this challenge addresses the real world setting, where paired true high and low-resolution images are unavailable. For training, only one set of unpaired source and target input images were provided to the participants. The source images have unknown degradations, while the target images are clean, high quality images. The challenge contains two tracks, where the goal was to super-resolve images with Image Processing artifacts (Track 1) or low-quality smart-phone images (Track 2). The challenge had in total 22 teams competing in the final step. Most of the participating were influenced AIM 2019 and demonstrated interesting and innovative solutions. Our goal is that this challenge stimulates future research in the area of unsupervised learning for image super-resolution and other similar tasks, by serving as a standard benchmark and by the establishment of new baseline methods.

Acknowledgements

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

Appendix A. Teams and affiliations

AIM2019 organizers

Members:

Andreas Lugmayr (andreas.lugmayr@vision.ee.ethz.ch)

Martin Danelljan (martin.danelljan@vision.ee.ethz.ch)

Radu Timofte (radu.timofte@vision.ee.ethz.ch)

Affiliation: Computer Vision Lab, ETH Zurich

SR-DL

Title: Generative Variational AutoEncoder for Real Image Super-Resolution

Team Leader:

Zhi-Song Liu (zhi-song.liu@inria.fr)

Members:

Zhi-Song Liu, LIX - Computer science laboratory at the cole polytechnique [Palaiseau]

Li-Wen Wang, Center of Multimedia Signal Processing, The Hong Kong Polytechnic University

Marie-Paule Cani, LIX - Computer science laboratory at the cole polytechnique [Palaiseau]

Wan-Chi Siu, Center of Multimedia Signal Processing, The Hong Kong Polytechnic University

MSMers

Title: Cycle-based Residual Channel Attention Network for Real-World Super-Resolution

Team Leader:

Fuzhi Yang (yfzcropy0702@sjtu.edu.cn)

Members:

Fuzhi Yang, Shanghai Jiao Tong University,

Huan Yang, Microsoft Research Beijing, P.R. China,

Jianlong Fu, Microsoft Research Beijing, P.R. China,

GDUT-wp

Title: Ensemble of ResNets for Image Restoration

Team Leader:

Hao Li (2111903004@mail2.gdut.edu.cn)

Members:

Hao Li, Guangdong University of Technology

Yukai Shi, Guangdong University of Technology

Junyang Chen, Guangdong University of Technology

KU-ISPL

Title: Unsupervised Real-World Super Resolution with Cycle-in-Cycle Generative Adversarial Network and Domain-Transfer Discriminator.

Team Leader:

Gwantae Kim (gtkim@ispl.korea.ac.kr)

Members:

Gwantae, Kim, Intelligent Signal Processing Laboratory, Korea University

Kanghyu, Lee, Intelligent Signal Processing Laboratory, Korea University

Jaihyun, Park, Intelligent Signal Processing Laboratory, Korea University

Junyeop, Lee, Intelligent Signal Processing Laboratory, Korea University

Jeongki, Min, Intelligent Signal Processing Laboratory, Korea University

Bokyeung, Lee, Intelligent Signal Processing Laboratory, Korea University

Hanseok, Ko, Intelligent Signal Processing Laboratory, Korea University

TeamAY

Title: SimUSR: A Simple but Strong Baseline for Unsupervised Image Super-resolution

Team Leader:

Namhyuk Ahn (aa0dfg@ajou.ac.kr)

Members:

Namhyuk, Ahn, Ajou University

Jaejun, Yoo, EPFL

Kyung-Ah, Sohn, Ajou University

MLP-SR

Title: Deep Generative Adversarial Residual Convolutional Networks for Real-World Super-Resolution

Team Leader:

Rao Muhammad Umer (engr.raoumer943@gmail.com)

Members:

Rao Muhammad Umer, University of Udine, Italy.

Christian Micheloni, University of Udine, Italy.

BOE-IOT-AIBD

Title: DSGAN and Triple-V MGBPv2 for Real Super-Resolution

Team Leader:

Pablo Navarrete Michelini (pnavarre@boe.com.cn)

Members:

Pablo, Navarrete Michelini, BOE Technology Group Co. Ltd.

Fengshuo, Hu, BOE Technology Group Co. Ltd.

Yanhong, Wang, BOE Technology Group Co. Ltd.

Yunhua, Lu, BOE Technology Group Co. Ltd.

SuperT

Title: Fast and Balanced Laplacian Pyramid Networks for Progressive image super-resolution

Team Leader:

Tongtong Zhao (daitoutiere@gmail.com)

Members:

Jinjia,Peng, Dalian Maritime University

Huibing,Wang, Dalian Maritime University

BIGFEATURE-CAMERA

Title: Deep Residual Mix Attention Network for Image Super-Resolution

Team Leader:

Kaihua Cheng (consonwm0909@gmail.com)

Members:

Kaihua Cheng, Guangdong OPPO Mobile Telecommunications Corp., Ltd

Haijie Zhuo, Guangdong OPPO Mobile Telecommunications Corp., Ltd

KU-ISPL2

Title: Modular generative adversarial network based super-resolution

Team Leader:

Kanghyu Lee (khlee@ispl.korea.ac.kr)

Members:

Gwantae Kim is with Department of Video Information Processing, Korea University

Junyeop Lee is with School of Electrical Engineering, Korea University

Jeongki Min is with School of Electrical Engineering, Korea University

Bokyeung Lee is with School of Electrical Engineering, Korea University

Jaihyun Park is with School of Electrical Engineering, Korea University

Hanseok Ko is with School of Electrical Engineering, Korea University

Impressionism

Title: Real World Super-Resolution via Kernel Estimation and Noise Injection

Team Leader:

Xiaozhong Ji (shawn_ji@163.com)

Members:

Xiaozhong Ji, Tencent YouTu Lab Yun Cao, Tencent YouTu

Lab Ying Tai, Tencent YouTu Lab Chengjie Wang, Tencent

YouTu Lab Jilin Li, Tencent YouTu Lab Feiyue Huang, Ten-

tencent YouTu Lab

Relbmag Eht

Title: Network of Aggregated Downampler-Upsampler with Dual Resolution Domain Matching

Team Leader:

Timothy Haoning Wu (1700012826@pku.edu.cn)

Members:

Haoning Wu, Peking University

ITS425

Title: Addptive Residual dense block network for real image super-resolution

Team Leader:

Ziyao Zong (824924664@qq.com)

Members:

Ziyao Zong, North China University of Technology

Shuai Liu, North China University of Technology

Biao Yang, North China University of Technology

AITA-Noah

Title: Real World Super-Resolution with Iterative Domain Adaptation, CycleSR and Conditional Frequency Separation GAN

Team Leader:

Ziluan Liu (liuziluan@huawei.com), Xueyi Zou (zouxueyi@huawei.com)

Members:

Xing Liu, Huawei Technologies Co., Ltd

Shuaijun Chen, Huawei Technologies Co., Ltd

Lei Zhao, Huawei Technologies Co., Ltd

Zhan Wang, Huawei Technologies Co., Ltd

Yuxuan Lin, Huawei Technologies Co., Ltd

Xu Jia, Huawei Technologies Co., Ltd

Ziluan Liu, Huawei Technologies Co., Ltd

Xueyi Zou, Huawei Technologies Co., Ltd

Webbzhou

Title: Guided Frequency Separation Network for Real World Super-Resolution

Team Leader:

Yuanbo Zhou (webbozhou@gmail.com)

Members:

Yuanbo Zhou, Fuzhou University Tong Tong, Fuzhou University, Imperial Vision Technology

Qinquan Gao, Fuzhou University, Imperial Vision Technology

Wei Deng, Imperial Vision Technology

Samsung-SLSI-MSL

Title: Real-World Super-Resolution using Generative Adversarial Networks

Team Leader:

Haoyu Ren, Amin Kheradmand (co-leader) (haoyu.ren@samsung.com)

Members:

Haoyu Ren SOC R&D, Samsung Semiconductor, Inc., USA

Amin Kheradmand SOC R&D, Samsung Semiconductor, Inc., USA

Mostafa El-Khamy SOC R&D, Samsung Semiconductor, Inc., USA

Shuangquan Wang SOC R&D, Samsung Semiconductor, Inc., USA

Dongwoon Bai SOC R&D, Samsung Semiconductor, Inc., USA

Jungwon Lee SOC R&D, Samsung Semiconductor, Inc.,
USA

BMIPL-UNIST-YH-1

Title: Unpaired Domain-Adaptive Image Super-Resolution using Cycle-Consistent Adversarial Networks

Team Leader:

YongHyeok Seo (syh4661@unist.ac.kr)

Members:

SeYoung Chun, Ulsan national institute of science and technology

SVNIT1

Title: Unsupervised Real-World Single Image Super-Resolution (SR) using Generative Adversarial Networks (GAN) and Variational Auto-Encoders (VAE)

Team Leader:

Kalpesh Prajapati (kalpesh.jp89@gmail.com)

Members:

Heena, Patel, Sardar Vallabhbhai National Institute Of Technology, Surat

Vishal, Chudasama, Sardar Vallabhbhai National Institute Of Technology, Surat

Kishor, Upla, Norwegian University of Science and Technology, Gjøvik, Norway

Raghavendra, Ramachandra, Norwegian University of Science and Technology, Gjøvik, Norway

Kiran, Raja, Norwegian University of Science and Technology, Gjøvik, Norway

Christoph, Busch, Norwegian University of Science and Technology, Gjøvik, Norway

SVNIT2

Title: Unsupervised Real-World Single Image Super-Resolution (SR)

Team Leader:

Kalpesh Prajapati (kalpesh.jp89@gmail.com)

Members:

Heena, Patel, Sardar Vallabhbhai National Institute Of Technology, Surat

Vishal, Chudasama, Sardar Vallabhbhai National Institute Of Technology, Surat

Kishor, Upla, Norwegian University of Science and Technology, Gjøvik, Norway

Raghavendra, Ramachandra, Norwegian University of Science and Technology, Gjøvik, Norway

Kiran, Raja, Norwegian University of Science and Technology, Gjøvik, Norway

Christoph, Busch, Norwegian University of Science and Technology, Gjøvik, Norway

InnoPeak-SR

Title: Deep Residual Channel Attention Generative Adversarial Networks for Image Super-Resolution and Noise Reduction

Team Leader:

Jie Cai (jie.cai@innopeaktech.com)

Members:

Jie Cai, InnoPeak Technology

Zibo Meng, InnoPeak Technology

Chiu Man Ho, InnoPeak Technology

References

- [1] Abdelrahman Abdelhamed, Mahmoud Afifi, Radu Timofte, Michael Brown, et al. NTIRE 2020 challenge on real image denoising: Dataset, methods and results. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2020. [2](#)
- [2] Namhyuk Ahn, Byungkon Kang, and Kyung-Ah Sohn. Fast, accurate, and lightweight super-resolution with cascading residual network. In *ECCV*, 2018. [1](#)
- [3] Namhyuk Ahn, Byungkon Kang, and Kyung-Ah Sohn. Image super-resolution via progressive cascading residual network. In *CVPR*, 2018. [1](#)
- [4] Codruta O. Ancuti, Cosmin Ancuti, Florin-Alexandru Vasluiianu, Radu Timofte, et al. NTIRE 2020 challenge on nonhomogeneous dehazing. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2020. [2](#)
- [5] Boaz Arad, Radu Timofte, Yi-Tun Lin, Graham Finlayson, Ohad Ben-Shahar, et al. NTIRE 2020 challenge on spectral reconstruction from an rgb image. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2020. [2](#)
- [6] Isabelle Begin and FR Ferrie. Blind super-resolution using a learning-based approach. In *ICPR*, 2004. [2](#)
- [7] Sefi Bell-Kligler, Assaf Shocher, and Michal Irani. Blind super-resolution kernel estimation using an internal-gan. In *NeurIPS*, pages 284–293, 2019. [3, 8](#)
- [8] Adrian Bulat, Jing Yang, and Georgios Tzimiropoulos. To learn image super-resolution, use a gan to learn how to do image degradation first. *arXiv preprint arXiv:1807.11458*, 2018. [2](#)
- [9] Jie Cai, Zibo Meng, and Chiu Man Ho. Residual channel attention generative adversarial network for image super-resolution and noise reduction. In *CVPR Workshops*, 2020. [11](#)
- [10] Shuaijun Chen, Enyan Dai, Zhen Han, Xu Jia, Ziluan Liu, Xing Liu, Xueyi Zou, Chunjing Xu, Jianzhuang Liu, and Qi Tian. Unsupervised image super-resolution with an indirect supervised path. In *CVPRW*, 2020. [8](#)
- [11] Yunjin Chen and Thomas Pock. Trainable nonlinear reaction diffusion: A flexible framework for fast and effective image restoration. *IEEE Trans. Pattern Anal. Mach. Intell.*, 39(6):1256–1272, 2017. [11](#)

- [12] Casey Chu, Andrey Zhmoginov, and Mark Sandler. Cy-clegan, a master of steganography. *arXiv preprint arXiv:1712.02950*, 2017. 5
- [13] Dengxin Dai, Radu Timofte, and Luc Van Gool. Jointly optimized regressors for image super-resolution. *Comput. Graph. Forum*, 34(2):95–104, 2015. 1
- [14] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Learning a deep convolutional network for image super-resolution. In *ECCV*, 2014. 1
- [15] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Image super-resolution using deep convolutional networks. *TPAMI*, 38(2):295–307, 2016. 1
- [16] Yuchen Fan, Honghui Shi, Jiahui Yu, Ding Liu, Wei Han, Haichao Yu, Zhangyang Wang, Xinchao Wang, and Thomas S Huang. Balanced two-stage residual networks for image super-resolution. In *CVPR*, 2017. 1
- [17] William T Freeman, Thouis R Jones, and Egon C Pasztor. Example-based super-resolution. *IEEE Computer graphics and Applications*, 2002. 1
- [18] Manuel Fritzsche, Shuhang Gu, and Radu Timofte. Frequency separation for real-world super-resolution. In *2019 IEEE/CVF International Conference on Computer Vision Workshops, ICCV Workshops 2019, Seoul, Korea (South), October 27-28, 2019*, pages 3599–3608, 2019. 3, 5, 10, 11, 12, 13
- [19] Dario Fuoli, Zhiwu Huang, Martin Danelljan, Radu Timofte, et al. NTIRE 2020 challenge on video quality mapping: Methods and results. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2020. 2
- [20] Jinjin Gu, Hannan Lu, Wangmeng Zuo, and Chao Dong. Blind super-resolution with iterative kernel correction. In *CVPR*, 2019. 2
- [21] Shuhang Gu, Martin Danelljan, Radu Timofte, et al. Aim 2019 challenge on image extreme super-resolution: Methods and results. In *ICCV Workshops*, 2019. 1, 10
- [22] Shuhang Gu, Andreas Lugmayr, Martin Danelljan, Manuel Fritzsche, Julien Lamour, and Radu Timofte. DIV8K: diverse 8k resolution image dataset. In *2019 IEEE/CVF International Conference on Computer Vision Workshops, ICCV Workshops 2019, Seoul, Korea (South), October 27-28, 2019*, pages 3512–3516, 2019. 9
- [23] Ishaan Gulrajani, Faruk Ahmed, Martín Arjovsky, Vincent Dumoulin, and Aaron C. Courville. Improved training of wasserstein gans. In *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, 4-9 December 2017, Long Beach, CA, USA*, pages 5767–5777, 2017. 10
- [24] Zhen Han, Enyan Dai, Xu Jia, Xiaoying Ren, Shuaijun Chen, Chunjing Xu, Jianzhuang Liu, and Qi Tian. Unsupervised image super-resolution with an indirect supervised path. *CoRR*, abs/1910.02593, 2019. 8
- [25] Mostafa El-Khamy Shuangquan Wang Dongwoon Bai Jung-won Lee Haoyu Ren, Amin Kheradmand. Real-world super-resolution using generative adversarial networks. In *CVPRW*, 2020. 10
- [26] Muhammad Haris, Gregory Shakhnarovich, and Norimichi Ukita. Deep back-projection networks for super-resolution. In *CVPR*, 2018. 1
- [27] Yingkun Hou, Chunxia Zhao, Deyun Yang, and Yong Cheng. Comments on "image denoising by sparse 3-d transform-domain collaborative filtering". *IEEE Trans. Image Processing*, 20(1):268–270, 2011. 12
- [28] Jia-Bin Huang, Abhishek Singh, and Narendra Ahuja. Single image super-resolution from transformed self-exemplars. In *CVPR*, 2015. 1
- [29] Yiwen Huang and Ming Qin. Densely connected high order residual network for single frame image super resolution. *arXiv preprint arXiv:1804.05902*, 2018. 1
- [30] Andrey Ignatov, Nikolay Kobyshev, Radu Timofte, Kenneth Vanhoey, and Luc Van Gool. Dslr-quality photos on mobile devices with deep convolutional networks. In *ICCV*, pages 3297–3305. IEEE Computer Society, 2017. 2, 6, 10
- [31] Andrey Ignatov, Nikolay Kobyshev, Radu Timofte, Kenneth Vanhoey, and Luc Van Gool. WESPE: weakly supervised photo enhancer for digital cameras. In *2018 IEEE Conference on Computer Vision and Pattern Recognition Workshops, CVPR Workshops 2018, Salt Lake City, UT, USA, June 18-22, 2018*, pages 691–700, 2018. 12
- [32] Andrey Ignatov, Radu Timofte, Thang Van Vu, Tung Minh Luu, Trung X Pham, Cao Van Nguyen, Yongwoo Kim, Jae-Seok Choi, Munchurl Kim, Jie Huang, et al. Pirm challenge on perceptual image enhancement on smartphones: Report. *arXiv preprint arXiv:1810.01641*, 2018. 4
- [33] Michal Irani and Shmuel Peleg. Improving resolution by image registration. *CVGIP*, 1991. 1
- [34] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. 2017. 8
- [35] Xiaozhong Ji, Yun Cao, Ying Tai, Chengjie Wang, Jilin Li, and Feiyue Huang. Real world super-resolution via kernel estimation and noise injection. In *CVPRW*, 2020. 8
- [36] Gwangtae Kim, Kanghyu Lee, Junyeop Lee, Jeongki Min, Bokyung Lee, Jaihyun Park, David K. Han, and Hanseok Ko. Unsupervised real-world super resolution with cycle generative adversarial network and domain discriminator. In *CVPR Workshops*, 2020. 11
- [37] Heewon Kim, Myungsub Choi, Bee Lim, and Kyoung Mu Lee. Task-aware image downscaling. *ECCV*, 2018. 2
- [38] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee. Accurate image super-resolution using very deep convolutional networks. In *CVPR*, 2016. 1
- [39] Wei-Sheng Lai, Jia-Bin Huang, Narendra Ahuja, and Ming-Hsuan Yang. Deep laplacian pyramid networks for fast and accurate super-resolution. In *CVPR*, pages 5835–5843. IEEE Computer Society, 2017. 13
- [40] Wei-Sheng Lai, Jia-Bin Huang, Narendra Ahuja, and Ming-Hsuan Yang. Deep laplacian pyramid networks for fast and accurate super-resolution. In *CVPR*, 2017. 1
- [41] Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew P Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, et al. Photo-realistic single image super-resolution using a generative adversarial network. *CVPR*, 2017. 1, 11, 13

- [42] Stamatios Lefkimiatis. Universal denoising networks : A novel CNN architecture for image denoising. In *2018 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2018, Salt Lake City, UT, USA, June 18-22, 2018*, pages 3204–3213, 2018. 11
- [43] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. Enhanced deep residual networks for single image super-resolution. *CVPR*, 2017. 1
- [44] Zhisong Liu, Wan-Chi Siu, Marie-Paule Cani, Li-Wen Wang, and Chu-Tak Li. Unsupervised real image super-resolution via generative variational autoencoder. In *CVPR Workshops*, 2020. 12
- [45] Andreas Lugmayr, Martin Danelljan, and Radu Timofte. Unsupervised learning for real-world super-resolution. In *ICCV Workshops*, 2019. 2, 3, 8, 10, 13
- [46] Andreas Lugmayr, Martin Danelljan, Radu Timofte, et al. Aim 2019 challenge on real-world image super-resolution: Methods and results. In *ICCV Workshops*, 2019. 2, 3, 5
- [47] Andreas Lugmayr, Martin Danelljan, Radu Timofte, et al. NTIRE 2020 challenge on real-world image super-resolution: Methods and results. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2020. 2
- [48] Chao Ma, Chih-Yuan Yang, Xiaokang Yang, and Ming-Hsuan Yang. Learning a no-reference quality metric for single-image super-resolution. *Comput. Vis. Image Underst.*, 158:1–16, 2017. 4
- [49] Tomer Michaeli and Michal Irani. Nonparametric blind super-resolution. In *ICCV*, 2013. 2
- [50] Pablo Navarrete Michelini, Wenbin Chen, Hanwen Liu, and Dan Zhu. Mgbpv2: Scaling up multi-grid back-projection networks. In *2019 IEEE/CVF International Conference on Computer Vision Workshops, ICCV Workshops 2019, Seoul, Korea (South), October 27-28, 2019*, pages 3399–3407, 2019. 10
- [51] A Mittal, AK Moorthy, and AC Bovik. Referenceless image spatial quality evaluation engine. In *45th Asilomar Conference on Signals, Systems and Computers*, volume 38, pages 53–54, 2011. 4
- [52] Anish Mittal, Rajiv Soundararajan, and Alan C. Bovik. Making a "completely blind" image quality analyzer. *IEEE Signal Process. Lett.*, 20(3):209–212, 2013. 4, 9
- [53] Venkatanath N., Praneeth D., Maruthi Chandrasekhar Bh., Sumohana S. Channappayya, and Swarup S. Medasani. Blind image quality evaluation using perception based features. In *NCC*, pages 1–6. IEEE, 2015. 4
- [54] Seungjun Nah, Sanghyun Son, Radu Timofte, Kyoung Mu Lee, et al. NTIRE 2020 challenge on image and video de-blurring. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2020. 2
- [55] Ahn Namhyuk, Jaejun Yoo, and Kyung-Ah Sohn. Simusr: A simple but strong baseline for unsupervised image super-resolution. In *CVPR Workshops*, 2020. 12
- [56] Sung Cheol Park, Min Kyu Park, and Moon Gi Kang. Super-resolution image reconstruction: a technical overview. *IEEE signal processing magazine*, 2003. 1
- [57] Kalpesh J Prajapati, Vishal Chudasama, Heena Patel, Kishor Upla, Raghavendra Ramachandra, Kiran Raja, and Christoph Busch. Unsupervised single image super-resolution network (usisresnet) for real-world data using generative adversarial network. In *CVPRW*, 2020. 13
- [58] Alec Radford, Luke Metz, and Soumith Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. In *4th International Conference on Learning Representations, ICLR 2016, San Juan, Puerto Rico, May 2-4, 2016, Conference Track Proceedings*, 2016. 13
- [59] Haoyu Ren, Mostafa El-Khamy, and Jungwon Lee. Dn-resnet: Efficient deep residual network for image denoising. In *Computer Vision - ACCV 2018 - 14th Asian Conference on Computer Vision, Perth, Australia, December 2-6, 2018, Revised Selected Papers, Part V*, pages 215–230, 2018. 9
- [60] Assaf Shocher, Nadav Cohen, and Michal Irani. Zero-shot super-resolution using deep internal learning. In *CVPR*, 2018. 12
- [61] Libin Sun and James Hays. Super-resolution from internet-scale scene matching. In *ICCP*, 2012. 1
- [62] Radu Timofte, Eirikur Agustsson, Luc Van Gool, Ming-Hsuan Yang, Lei Zhang, Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, Kyoung Mu Lee, et al. Ntire 2017 challenge on single image super-resolution: Methods and results. *CVPR Workshops*, 2017. 2
- [63] Radu Timofte, Vincent De Smet, and Luc Van Gool. A+: Adjusted anchored neighborhood regression for fast super-resolution. In *ACCV*, pages 111–126. Springer, 2014. 1
- [64] Radu Timofte, Rasmus Rothe, and Luc Van Gool. Seven ways to improve example-based single image super resolution. In *CVPR*, pages 1865–1873. IEEE Computer Society, 2016. 1, 13
- [65] Radu Timofte, Vincent De Smet, and Luc Van Gool. Anchored neighborhood regression for fast example-based super-resolution. In *ICCV*, pages 1920–1927, 2013. 1
- [66] Rao Muhammad Umer, Gian Luca Foresti, and Christian Micheloni. Deep generative adversarial residual convolutional networks for real-world super-resolution. In *CVPRW*, 2020. 13
- [67] Xintao Wang, Ke Yu, Shixiang Wu, Jinjin Gu, Yihao Liu, Chao Dong, Chen Change Loy, Yu Qiao, and Xiaou Tang. Esgan: Enhanced super-resolution generative adversarial networks. *ECCV*, 2018. 1, 2, 3, 8, 11, 12, 13
- [68] Zhou Wang, Alan C. Bovik, Hamid R. Sheikh, and Eero P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Image Processing*, 13(4):600–612, 2004. 3
- [69] Chih-Yuan Yang and Ming-Hsuan Yang. Fast direct super-resolution by simple functions. In *ICCV*, pages 561–568, 2013. 1
- [70] Jianchao Yang, John Wright, Thomas S. Huang, and Yi Ma. Image super-resolution as sparse representation of raw image patches. In *CVPR*, 2008. 1
- [71] Jianchao Yang, John Wright, Thomas S. Huang, and Yi Ma. Image super-resolution via sparse representation. *IEEE Trans. Image Processing*, 19(11):2861–2873, 2010. 1

- [72] Xin Yu and Fatih Porikli. Ultra-resolving face images by discriminative generative networks. In *ECCV*, pages 318–333, 2016. 1
- [73] Shixin Yuan, Radu Timofte, Ales Leonardis, Gregory Slabaugh, et al. NTIRE 2020 challenge on image demoiréing: Methods and results. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2020. 2
- [74] Yuan Yuan, Siyuan Liu, Jiawei Zhang, Yongbing Zhang, Chao Dong, and Liang Lin. Unsupervised image super-resolution using cycle-in-cycle generative adversarial networks. *CVPR Workshops*, 2018. 2, 11
- [75] Kai Zhang, Shuhang Gu, Radu Timofte, et al. NTIRE 2020 challenge on perceptual extreme super-resolution: Methods and results. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2020. 2
- [76] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. *CVPR*, 2018. 3, 9
- [77] Yulun Zhang, Kunpeng Li, Kai Li, Lichen Wang, Bineng Zhong, and Yun Fu. Image super-resolution using very deep residual channel attention networks. In *Computer Vision - ECCV 2018 - 15th European Conference, Munich, Germany, September 8-14, 2018, Proceedings, Part VII*, pages 294–310, 2018. 3, 9, 10, 12, 13
- [78] Yulun Zhang, Kunpeng Li, Kai Li, Bineng Zhong, and Yun Fu. Residual non-local attention networks for image restoration. In *7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019*, 2019. 13
- [79] Yulun Zhang, Yapeng Tian, Yu Kong, Bineng Zhong, and Yun Fu. Residual dense network for image super-resolution. In *2018 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2018, Salt Lake City, UT, USA, June 18-22, 2018*, pages 2472–2481, 2018. 11
- [80] Yuanbo Zhou, Wei Deng, Tong Tong, and Qinquan Gao. Guided frequency separation network for real-world super-resolution. In *CVPR Workshops*, 2020. 12
- [81] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. *ICCV*, 2017. 9, 10