NTIRE 2022 Challenge on Learning the Super-Resolution Space

Andreas Lugmayr* Martin Danelljan* Radu Timofte* Kang-wook Kim
Younggeun Kim Jae-young Lee Zechao Li Jinshan Pan Dongseok Shim
Ki-Ung Song Jinhui Tang Cong Wang Zhihao Zhao

Abstract

This paper reviews the NTIRE 2022 challenge on learning the super-Resolution space. This challenge aims to raise awareness that the super-resolution problem is ill-posed. Since many high-resolution images map to the same low-resolution image, we asked the participants to create methods that sample diverse super-resolution from the space of possible high-resolution images given a low-resolution image. For evaluation, we use the same protocol as introduced in the last year's super-resolution space challenge of NTIRE 2021. We compare the submissions of the participating teams and relate them to the approaches from last year. This challenge contains two tracks: $4\times$ and $8\times$ scale factor. In total, 3 teams competed in the final testing phase.

1. Introduction

The generation of a high-resolution conditioned on a low-resolution image is called super-resolution. This discipline has a long tradition in computer vision [25, 17, 48, 56, 54, 60, 61, 62, 53, 10, 23, 55, 13, 14, 28, 31, 36, 16, 2, 3, 20, 24, 19] and is used in many different subfields. With the performance leap of deep learning also most super-resolution methods adopted this technology. As loss for those learned methods early works optimized for pixel-wise errors, which are favorable for the PSNR metric [13, 14, 28, 31, 36]. However, optimizing for this metric results in blurry textures. This was addressed in the works [64, 32, 58] using adversarial losses.

Since super-resolution is often formulated as inverting a downsampling kernel, which removes information, it is an ill-posed problem. Therefore, infinitely many high-resolution images can be mapped to the same low-resolution image. Inversion is hence ambiguous, and to explore the space of plausible high-resolution images given a low-resolution image, methods need to be stochastic. To sample from the distribution of high-resolution images, GAN-based approaches were adapted to generate a stochastic output [6, 8]. Furthermore, conditional Flow-based methods were found to generate rich diversity for super-resolution [41, 42, 59, 35, 26, 29]. Further technologies used to generate diverse images use VAEs [37, 66] and IMLE [33].

In recent years stochastic generative method using Diffusion Models [52] gained in performance [21, 22, 46, 11]. They were made conditional for various tasks like color-guided image-generation, super-resolution, inpainting and colorization [9, 43, 38, 51, 47].

The advantages of methods that can sample super-resolutions given a low-resolution image are that one can choose from multiple predictions or use it to steer the super-resolution with further conditioning.

The NTIRE 2022 Learning the Super-resolution Space challenge aims to deepen research in the direction of stochastic super-resolution and improve the state-of-the-art of SR in general. The participants are evaluated using the same metrics as in last year's challenge [39], which are photo-realism, consistency with the LR image, and how well the SR space is spanned.

This challenge is one of the NTIRE 2022 associated challenges: spectral recovery [5], spectral demosaicing [4], perceptual image quality assessment [18], inpainting [50], night photography rendering [15], efficient super-resolution [34], learning the super-resolution space [40], super-resolution and quality enhancement of compressed video [63], high dynamic range [49], stereo super-resolution [57], burst super-resolution [7].

2. NTIRE 2022 Challenge

The goals of the NTIRE 2022 Learning the Super-Resolution Space Challenge is to (i) stimulate research into learning the full space of plausible super-resolution; (ii) establish a benchmark protocols and metrics for stochastic super-resolution; (iii) probe the state-of-the-art in super-resolution in general.

*Andreas Lugmayr (andreas.lugmayr@vision.ee.ethz.ch), Martin Danelljan, and Radu Timofte are the NTIRE 2022 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/ntire22/

Team	Generative formulation					Additional Data
	Flow	GAN	VAE	IMLE	Diffusion	
Deepest	✓					✓
IMAG_WZ					✓	✓
IMAG_ZW					✓	✓
Deepest (21)	✓					✓
FutureReference				✓		✓
SR_DL		✓	✓			
SSS		✓				✓

Table 1. Information about the participating teams in the challenge.

2.1. Overview

The challenge contains two tracks, targeting $4\times$ and $8\times$ super-resolution respectively. Evaluation code and information about the challenge were provided at the public GitHub page `http://git.io/SR22`. The challenge uses the train, validation and testing sets as defined in employs the DIV2k [1]. As the final result, the participants in the challenge were asked to submit 10 super-resolution samples for each given LR image.

2.2. Rules

To guide the research towards useful and generalizable techniques, submissions needed to adhere to the following rules.

- The method must be able to generate an arbitrary number of diverse samples. That is, the method cannot be limited to a maximum number of different SR samples (corresponding to *e.g.* a certain number of different output network heads).
- All SR samples must be generated by a single model. That is, no ensembles are allowed.
- No self-ensembles or test-time data augmentation (flipping, rotation, etc.).
- All SR samples must be generated using the same hyper-parameters. That is, the generated SR samples shall not be the result of different choices of hyper-parameters during inference.
- Submissions of deterministic methods were allowed. However, they will naturally score zero in the diversity measure and therefore not be able to win the challenge.
- Other than the validation and test split of the DIV2k dataset, any training data or pre-training is allowed.

Furthermore, all participants were asked to submit the code of their solution along with the final results.

2.3. Challenge phases

The challenge started on the 7th of February 2022 with providing the task, evaluation scripts, training, and evaluation data to the participating teams. After the teams developed their methods, they received the test input data on the 23rd of March 2022 and submitted their ten predictions per LR image, description, code, and models until the 30th of March 2022.

2.4. Data

We provide the standard DIV2K dataset for $4\times$ and $8\times$ for training and validation. For testing, we only provide the LR images of the test set for both Tracks.

3. Evaluation Protocol

A method is evaluated by first predicting a set of 10 randomly sampled SR images for each low-resolution image in the dataset. Evaluating metrics corresponding to the three criteria above will be considered from this set of images. First, the participating methods are ranked according to each metric. Secondly, these ranks are combined into a final score. The individual metrics are described below.

3.1. Photo-realism

Computing the perceived difference between two images as humans perceive them is challenging. As an approximation, we included the LPIPS distance [65] in the evaluation script that was provided along with the validation set. For the final measurement of photo-realism, we conducted a user study to determine the most realistic super-resolution when seeing the low-resolution image as perceived by humans.

User Study We use the same user-study protocol as last year, where we designed it in a way that directly measures the rank of all methods. Therefore, we created a web interface for the users where they can drag and drop crops of images into a ranked list according to their perceptual quality. We hence calculate the final rank by directly applying the mean over all user answers, resulting in the Mean Opinion Rank (MOR). Consistent with last year's evaluation, we use three 80×80 crops for each of the 100 DIV2K test images and ask five different users per crop.

3.2. The spanning of the SR Space

To evaluate how well a method spans the SR Space, we measure the diversity of each method. To measure it in a meaningful way and reduce the potential for adversarial attacks we use the robust diversity score introduced in the preceding challenge at NTIRE 2021 [39]. To rehearse the motivation of this score, we consider two cohorts of images, first texture-rich like fur, and second flat regions like a patch of sky.

In a texture-rich image, many high-resolution images are downsampled to the same low-resolution image spanning a conditional distribution. We consider this distribution as the space of plausible super-resolutions. Since the ground truth is an element in that distribution and the ideal super-resolution algorithm samples the whole space, one obtain an image which is arbitrarily close to the ground truth by sampling enough images. Therefore, methods that span the space of plausible super-resolutions gain closeness to the ground truth when sampling more super-resolutions.

In the case of uniform regions like sky, the super-resolution methods should not generate a diverse set of high-resolutions, but only the uniform patch. Methods that artificially add diversity for such regions generate structures that are not contained in the original image. Therefore such attempts do not improve in diversity score with increasing number of samples and further harms the method for the perceptual metric described above.

Another aspect being considered is the high-dimensionality of the high-resolution images. Since we consider mega-pixel images, the closeness to the ground truth on the entire image is almost the same for all samples of stochastic super-resolution methods. Different samples will have regions that are closer to the ground truth, and other regions that are closer to another plausible high-resolution image. Due to the high-dimensionality this effect evens out for the average distance to the ground truth when considering multiple samples. Therefore we consider local patches to measure the diversity score.

The used diversity score is as follows, where M is the number of sampled super-resolutions, y_k the k-th patch in the original HR image y, and $\hat{y}^i{}_{i=1}^M$ the samples from the super-resolution method. The detailed derivation can be found in [39].

$$S_M = \frac{1}{\bar{d}_M}\left(\bar{d}_M - \frac{1}{K}\sum_{k=1}^{K}\min\left\{d(y_k, \hat{y}_k^i)\right\}_{i=1}^{M}\right). \quad (1)$$

3.3. Low Resolution Consistency

To measure how much information is preserved in the super-resolved image from the low-resolution image, we measure the LR-PSNR. It is computed as the PSNR between the input LR image and the predicted sample down-sampled with the given bicubic kernel. The goal of this challenge is to obtain an LR-PSNR of at least 45dB.

4. 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. Three teams of the 54 registered participants submitted to the final test phase. The methods of the teams that entered the final phase are described in Section 5 and the teams' members and affiliations are shown in Section Appendix A.

4.1. Baselines

As in the first challenge [39], we compare the submitted method from this and last year to the following baselines.

ESRGAN To compare the submissions with a photo-realistic super-resolution method, we use ESRGAN [58] as reference. Since it is deterministic, the diversity score is zero.

SRFlow As the baseline with diverse super-resolution output, we use the Flow-based method SRFlow [41]. It conditions the image generation method [30] for super-resolution. Different from the ESRGAN, the generated super-resolution images are highly consistent with the input low-resolution image.

4.2. Architectures and Main Ideas

Here we discuss the main ideas of this and last year's submitted methods. The underlying technologies and the use of external data are indicated in Table 1.

Flow-Based The winning team "Deepest (21)" based their approach on SRFlow [41] and submitted a modified version this year. Their strategy is to train a Normalizing Flow model to transform a high-resolution image conditioned on a low-resolution image into a latent variable. The training objective is to minimize the negative log-likelihood of this latent variable belonging to the gaussian distribution. For inference, they use the property of Normalizing Flows [12] that they are invertible. They sample a latent vector of Gaussian noise and transform it, conditioned on the low-resolution image to a high-resolution image. Since the method is bijective, it cannot map two different latent

Team	LPIPS	LR-PSNR	Div. Score S_{10} [%]	MOR	Final Rank
IMAG_ZW	0.171	48.14	$21.938_{(3)}$	$3.57_{(2)}$	2.5
Deepest	0.126	50.13	$28.853_{(1)}$	$3.67_{(3)}$	2.5
IMAG_WZ	0.169	45.20	$27.320_{(2)}$	$3.34_{(1)}$	1.5
FutureReference (IMLE)	0.165	37.51	19.636	-	-
SR_DL (VAE)	0.234	39.80	20.508	-	-
SSS (GAN)	0.110	44.70	13.285	-	-
Deepest (Flow)	0.117	50.54	26.041	-	-
SRFlow	0.122	49.86	25.008	3.62	-
ESRGAN	0.124	38.74	0.000	3.52	-
GT	0	∞	-	3.15	-

Table 2. Quantitative comparison of participating teams. (4×)

Team	LPIPS	LR-PSNR	Div. Score S_{10} [%]	MOR	Final Rank
Deepest	0.257	50.37	26.539	4.510	-
FutureReference (IMLR)	0.291	36.51	17.985	4.741	-
SSS (GAN)	0.237	37.43	13.548	4.850	-
SR_DL (VAE-GAN)	0.311	42.28	14.817	4.797	-
Deepest (Flow)	0.259	48.64	26.941	4.503	-
SRFlow	0.282	47.72	25.582	4.775	-
ESRGAN	0.284	30.65	0	4.452	-
GT	0	∞	-	3.173	-

Table 3. Quantitative comparison of participating teams. (8×)

Figure 1. Qualitative comparison between the participating approaches for $4\times$ super-resolution

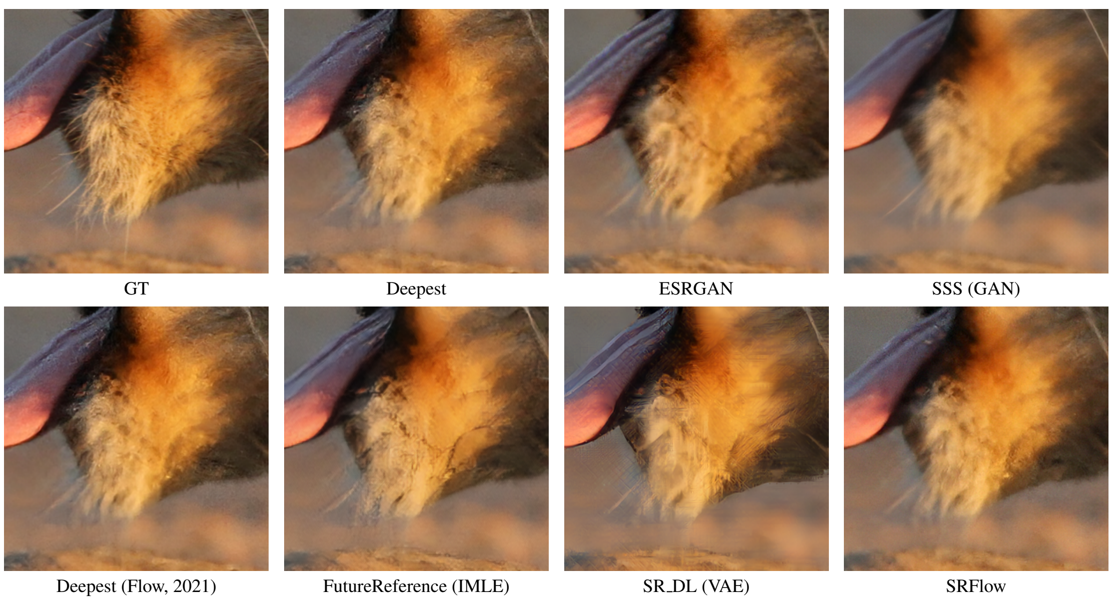

Figure 2. Qualitative comparison between the participating approaches for $8\times$ super-resolution

vectors to the same high-resolution image. Therefore, it creates a diverse set of super-resolutions. Furthermore, differently from ESRGAN, it was shown that SRFlow creates super-resolution that are consistent with the input. They downsampled the super-resolution using the same kernel as for the training pair generation and measured the PSNR between those two low-resolution images to show this property. The team Deepest worked on the information content gap between the HR image and the latent space and adopted frequency separation in this year's submission. Noteworthy, the team njtech&seu used multi-head attention in their approach to NTIRE 21 [39] and scored the highest Diversity Score in both $4\times$ and $8\times$. However, their perceptual quality did not suffice to outperform the baseline SRFlow.

Diffusion-Based The teams IMAG_ZW and IMAG_WZ submitted methods using diffusion models [21] which are known to produce highly stochastic output.

GAN-Based The best performing team in terms of diversity and perception of methods that relied on GAN-based approaches was the team SSS. We observed that GAN approaches struggled to generate large diversity in their super-resolutions. Furthermore, the adversarial loss encourages hallucinations in the super-resolution and therefore reduces the LR-PSNR. This method did not reach the set threshold of 45dB and was not considered for the final human study.

VAE-Based Of methods using VAEs for super-resolution,

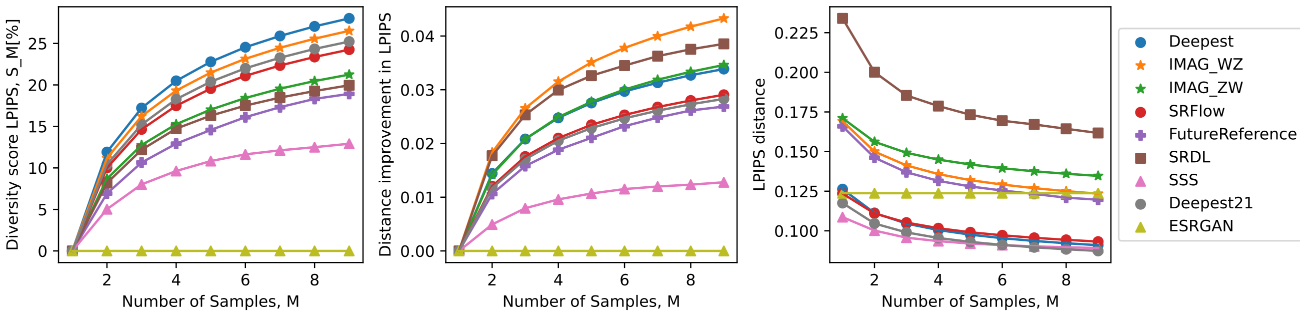

Figure 3. Visualization of improvement in LPIPS for $4\times$ by number of samples. Flow: Circle, VAE: Square, IMLE: Plus, GAN: Triangle, Diffusion: Star

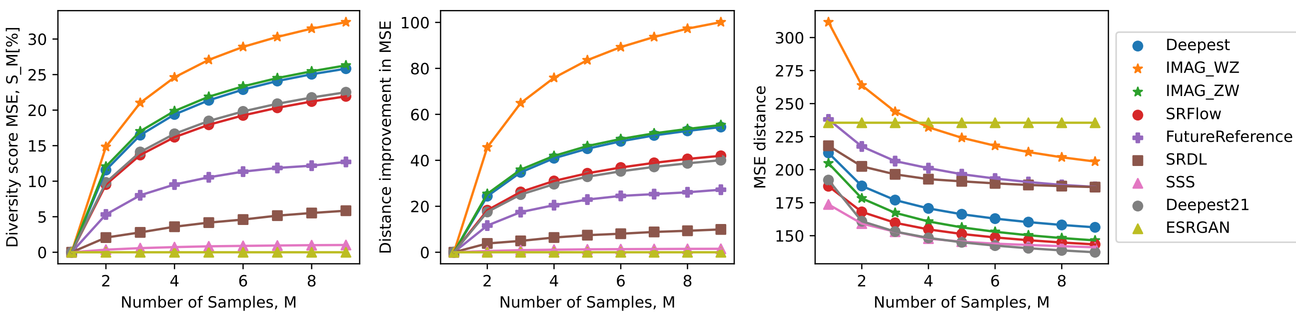

Figure 4. Visualization of improvement in MSE for $4\times$ by number of samples. Flow: Circle, VAE: Square, IMLE: Plus, GAN: Triangle, Diffusion: Star

the team SR_DL performed best. They leveraged the stochastic properties of VAEs to generate diverse outputs. An advantage of VAEs over Flow-Based methods is that they do not pose such strict architectural constraints as the bijectiveness and the tractability of the Jacobian. To further improve photo-realism, this team also employed an adversarial loss.

IMLE-based The team FutureReference, who was the only method using IMLE-based [44] super-resolution, showed that this approach is also capable of producing high-quality and diverse outputs. Their training objective reverses the generation process to match the super-resolutions with real data.

4.3. Discussion

In this section, we discuss the final evaluation results on the DIV2K test set of both the $4\times$ and $8\times$ super-resolution tracks. The quantitative results of this the three teams that participated this year, the top-performing teams for Flow, GAN, IMLE, and VAE in the middle section, and finally, the baselines are provided in Tables 2 and 3. The final evaluation for photo-realism is done employing the MOR described above. Since the number of methods that can be compared in the user-study is limited, we did not compare with all methods for which we report numerical results. For $4\times$, we calculated the score of this year's submissions and the baselines, and for $8\times$, we compared the method Deepest of this year with the previous year's methods and the baselines. Further, we report the LPIPS, LR-PSNR, and the diversity score. The final ranking is obtained by the average rank of the MOR and the diversity score.

The method with the highest perceptual quality measured by the MOR for $4\times$ was submitted by the team IMAG_WZ. They use diffusion models, which are known for their high perceptual quality. The superiority of this method can also be observed in the Figure 1. Further visual results for this method on $4\times$ are shown in Figure 9. Although diffusion models have a strong prior and can therefore be guided with little information, this team did not submit results for $8\times$.

The highest diversity score on $4\times$ was achieved by the team Deepest. They increased the score over last year's approach by using frequency separation. As all previously submitted purely Flow-based methods, this method achieves almost perfect consistency with the low-resolution image of $50.12dB$. Samples of super-resolution images from the same low-resolution patch can be seen in figure 10.

Higher diversity scores compared to the baseline SRFlow were only achieved by the Flow-Based and Diffusion Model-based methods submitted to this and last year's challenge. Furthermore, the GAN, VAE, and IMLE-based methods that we compare with achieved a low input consis-

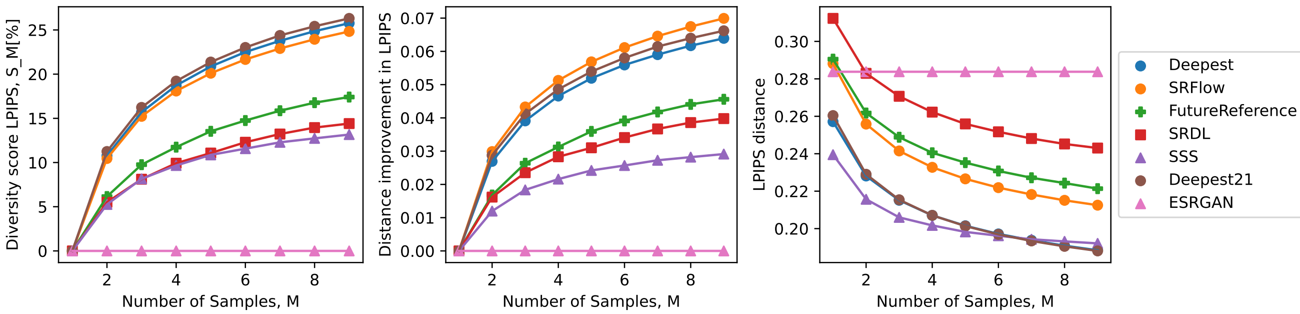

Figure 5. Visualization of improvement in LPIPS for $8\times$ by number of samples. Flow: Circle, VAE: Square, IMLE: Plus, GAN: Triangle

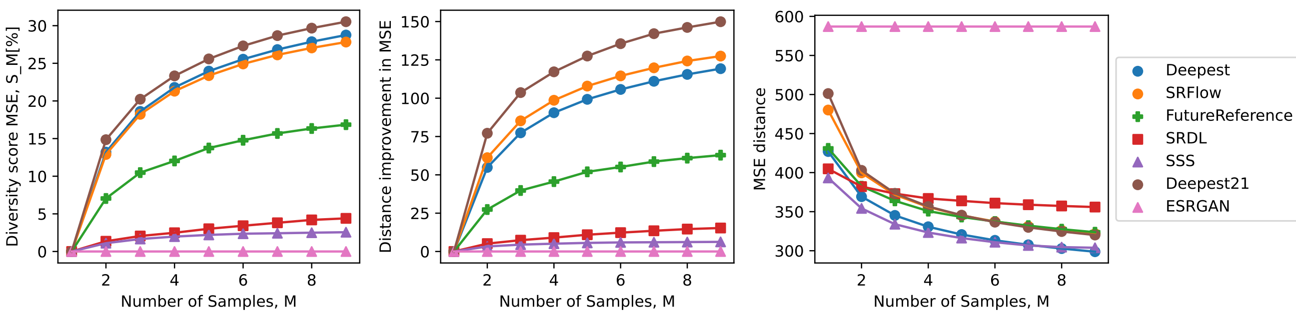

Figure 6. Visualization of improvement in MSE for $8\times$ by number of samples. Flow: Circle, VAE: Square, IMLE: Plus, GAN: Triangle

tency measured by the LR-PSNR, which was below 45dB. Note that the VAE-based method also uses an adversarial loss, which is known to lower the LR-PSNR. All three methods submitted this year achieved an LR-PSNR above 45db in both $4\times$ and $8\times$, which was set as a minimum to be considered low-resolution consistent.

To visualize the diversity score and two of its components, we plot it for all methods. Since the score is based on the assumption that an ideal super-resolution method can reach the GT arbitrary close, the diversity score should improve with an increasing number of samples. This behavior can be seen in Figures 3 and 5. Furthermore, the diversity score defined in NTIRE 21 [39] is not bound to a specific distance metric. Although we use the LPIPS as the primary metric, we show that the mean square error shows similar behavior 4 and 6. For this experiment, we smoothed the error maps with a moving average kernel of size 16, consistent with the metrics of last year.

Furthermore, we show two components of the score individually. On the right side, we depict $\min\left\{d(y_k, \hat{y}_k^i)\right\}_{i=1}^{M}$ showing the local minimum distance between the super-resolutions and the ground truth. In the middle, we plot $\bar{d}_M - \frac{1}{K}\sum_{k=1}^{K}\min\left\{d(y_k, \hat{y}_k^i)\right\}_{i=1}^{M}$, showing the absolute local impovement. On the left side we show the final diversity score, which divides the absolute improvement by the distance of the super-resolution which is globally closest to the ground truth. Without this last operation, the diversity score would favor methods with low distance, since it is harder to get closer to the ground truth if the best sample is already close.

This year the diffusion model-based IMAG_WZ overtook the Flow-based methods in the final score for $4\times$. Although the team Deepest, the improved version of last year's winning team, has a better diversity score, they have worse perceptual quality. For $8\times$ we only received one submission which achieves comparable results to last year's submissions. All participating methods outperform last year's approaches using VAE, GAN and IMLE by a large margin in diversity score.

5. Teams

5.1. Deepest

FS-NCSR: Increasing Diversity of Super-Resolution Space via Frequency Separation and Noise-Conditioned Normalizing Flow

This team proposes FS-NCSR (Frequency Separating Noise-Conditioned Normalizing Flow for Super-Resolution) where the generative model for super-resolution only produces the high-frequency elements of the target high-resolution image x without redundant low-frequency information. The low-frequency elements of the high-resolution input x are filtered out by the low-pass filter during training and the generative super-resolution architecture aims to estimate the high-frequency elements of the target. They utilize bicubic downsampling-upsampling

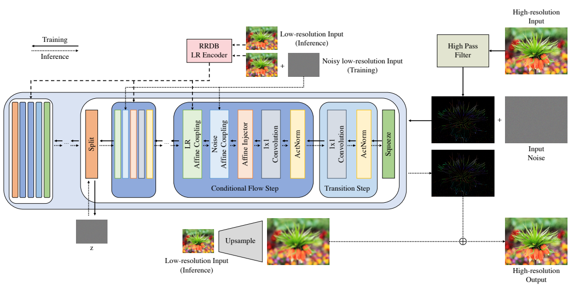

Figure 7. Algorithm overview. Our method proposes a frequency separation on the target image and applies noise on high-frequency input with noise-conditioned coupling layers for diverse super-resolution outputs.

as the low-pass filter with a specific scale factor s and the high-frequency input x_{hf} is calculated by subtracting low-frequency elements from the high-resolution target x,

$$x_{hf} = x - ((x)_{s\downarrow})_{s\uparrow}, \tag{2}$$

where $(\cdot)_{s\downarrow}$ and $(\cdot)_{s\uparrow}$ indicate bicubic downsampling and upsampling with the scale factor s respectively.

They leverage a normalizing flow based super-resolution model, SRFlow [41], as a baseline for high-quality diverse outputs compared to GAN which produces deterministic single outputs. The structure of our model basically follows SRFlow which consists of a squeeze module, transition step, conditional flow step, and split module. The main difference is that the input of the forward process in the normalizing flow is not the high-resolution image x as suggested in SRFlow, but the filtered high-frequency information x_{hf} of the high-resolution image.

$$x_{hf} = f_n \circ f_{n-1} \circ \cdots \circ f_1(z) \tag{3}$$

where z and $f(\cdot)$ indicate Gaussian latent variable and flow model which consists of invertible transformation.

The motivation of the flow-based architecture is to map the simple distribution p_z to the complex image distribution p_x with multi-layer invertible transformation. However, the mismatch of the manifold input and output data distribution induces poor generation performance, and SoftFlow [27] alleviate such mismatch by estimating a conditional perturbed data distribution rather than estimating direct input distribution.

NCSR [29] applies noise-conditioned affine coupling suggested in SoftFlow to the SRFlow architecture for diverse outputs without noisy artifacts. The transition step of our proposed method consists of 5 components: ActNorm, 1×1 convolution, affine injector, and two conditional affine coupling (noise affine coupling and low-resolution affine coupling), which are the same as NCSR. Gaussian noise is added to the input during the forward process in training,

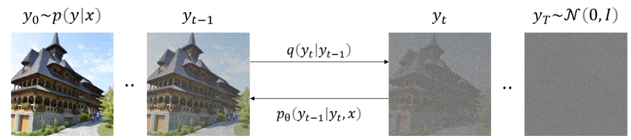

Figure 8. The forward diffusion process q(left to right) gradually adds noise to the target image. The reverse inference process p (right to left) The reverse process p is to restore the image under the conditions of the source image x by iterative method. Source image x is not shown here.

and our method applies noise on the high-frequency input x_{hf}.

$$\begin{aligned} v &\sim \mathcal{N}(0, \Sigma) \\ x_{hf}^{+} &= x_{hf} + v \\ y^{+} &= y + w \\ z &= f^{-1}(x_{hf}^{+}|y^{+}, v) \end{aligned} \tag{4}$$

where w indicates noise resized to the same size as the low-resolution input y.

They formulate the loss function only with negative log-likelihood $\mathcal{L}_{nll}$ similar to [41, 29],

$$\begin{aligned} \mathcal{L}_{nll} &= -\log p_{x|y,v}(x|y, v, \theta) \\ &= -\log p_z(f_\theta(x; y, v)) - \log|\det \frac{\partial f_\theta}{\partial y}(x; y, v)|. \end{aligned} \tag{5}$$

5.2. IMAG_WZ: Diffusion Models for Learning the Super Resolution Space and IMAG_ZW: Learning the Super-Resolution Space Using Diffusion Gamma Models

This team uses Conditional DDPMs, which generates a target image by T-step refinement. The model starts with a pure noise image, iteratively refines the corresponding image over T successful iterations according to the learned conditional transformation distribution. (see Figure 8)

The distribution of intermediate images in the inference chain is determined in the forward process that gradually adding noise to the signal through the Markov chain. The goal of our model is to reverse the diffusion process by iteratively recovering its target image from noise given an image through a reverse Markov chain.

As a result, they learn the reverse chain by using a neural denoising model which can estimate the noise.

5.2.1 IMAG_ZW: Learning the Super-Resolution Space Using Diffusion Gamma Models

This team submitted a variation of the approach from IMAG_WZ differing in the noise generation process. In the previous work on diffusion models, most of the methods

Figure 9. Visual example of diversity in super-resolution samples. The top left image is the input LR image, to the right is the ground truth and the ten remaining the samples from IMAG_WZ. ($4\times$)

are based on Gaussian noise, but some recent work [45] has shown us that the Gamma distribution can be better adapted to the estimated residual noise in the generation process. Moreover, these methods can also achieve competitive results in the generation process. Therefore, they introduced Gamma noise instead of Gaussian noise and trained the model with good results.

Acknowledgments

We thank the NTIRE 2022 sponsors: Huawei, Reality Labs, Bending Spoons, MediaTek, OPPO, Oddity, Voyage81, ETH Zurich (Computer Vision Lab) and University of Wurzburg (CAIDAS).

Appendix A. Teams and affiliations

NTIRE 2022 challenge organizers

Members:
Andreas Lugmayr (andreas.lugmayr@vision.ee.ethz.ch), Computer Vision Lab, ETH Zurich
Martin Danelljan (martin.danelljan@vision.ee.ethz.ch), Computer Vision Lab, ETH Zurich
Radu Timofte (radu.timofte@uni-wuerzburg.de), Computer Vision Lab, ETH Zurich/University of Würzburg

Deepest

Title: FS-NCSR: Increasing Diversity of Super-Resolution Space via Frequency Separation and Noise-Conditioned Normalizing Flow
Team Leader: Ki-Ung, Song
Members:
Ki-Ung, Song, Seoul National University

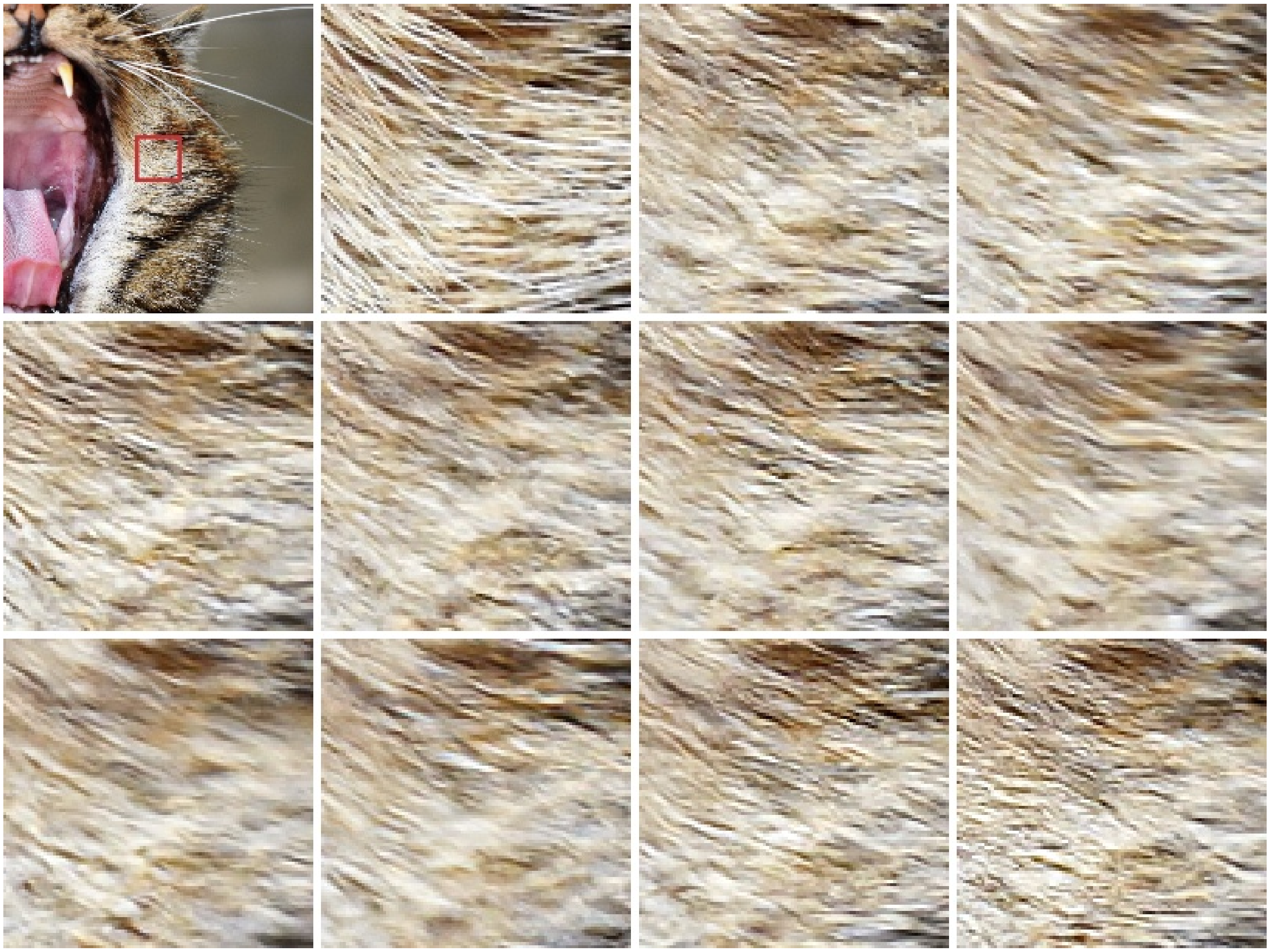

Figure 10. Visual example of diversity in super-resolution samples. The top left image is the input LR image, to the right is the ground truth and the ten remaining the samples from Deepest. (8×)

Dongseok, Shim, Seoul National University
Kang-wook, Kim, Seoul National University/Supertone Inc
Jae-young, Lee, Seoul National University
Younggeun, Kim, Seoul National University/MINDsLab Inc

IMAG_WZ

Title: Diffusion Models for Learning the Super Resolution Space

Team Leader: Cong, Wang

Members:
Zhihao, Zhao
Jinhui, Tang
Zechao, Li
Jinshan, Pan

Affiliation: Nanjing University of Science and Technology

IMAG_ZW

Title: Learning the Super-Resolution Space Using Diffusion Gamma Models

Team Leader: Zhihao, Zhao
Members:
Cong, Wang
Jinhui, Tang
Zechao, Li
Jinshan, Pan
Affiliation: Nanjing University of Science and Technology

References

[1] Eirikur Agustsson and Radu Timofte. Ntire 2017 challenge on single image super-resolution: Dataset and study. In *CVPR Workshops*, 2017.

[2] Namhyuk Ahn, Byungkon Kang, and Kyung-Ah Sohn. Fast, accurate, and lightweight super-resolution with cascading residual network. In *ECCV*, 2018.

[3] Namhyuk Ahn, Byungkon Kang, and Kyung-Ah Sohn. Image super-resolution via progressive cascading residual network. In *CVPR*, 2018.

[4] Boaz Arad, Radu Timofte, Rony Yahel, Nimrod Morag, Amir Bernat, et al. NTIRE 2022 spectral demosaicing challenge and dataset. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2022.

[5] Boaz Arad, Radu Timofte, Rony Yahel, Nimrod Morag, Amir Bernat, et al. NTIRE 2022 spectral recovery challenge and dataset. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2022.

[6] Yuval Bahat and Tomer Michaeli. Explorable super resolution. In *CVPR*, pages 2713–2722. IEEE, 2020.

[7] Goutam Bhat, Martin Danelljan, Radu Timofte, et al. NTIRE 2022 burst super-resolution challenge. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2022.

[8] Marcel C. Bühler, Andrés Romero, and Radu Timofte. Deepsee: Deep disentangled semantic explorative extreme super-resolution. In *ACCV*, volume 12625 of *Lecture Notes in Computer Science*, pages 624–642. Springer, 2020.

[9] Jooyoung Choi, Sungwon Kim, Yonghyun Jeong, Youngjune Gwon, and Sungroh Yoon. Ilvr: Conditioning method for denoising diffusion probabilistic models. *arXiv preprint arXiv:2108.02938*, 2021.

[10] Dengxin Dai, Radu Timofte, and Luc Van Gool. Jointly optimized regressors for image super-resolution. *Comput. Graph. Forum*, 34(2):95–104, 2015.

[11] Prafulla Dhariwal and Alex Nichol. Diffusion models beat gans on image synthesis. *arXiv preprint arXiv:2105.05233*, 2021.

[12] Laurent Dinh, David Krueger, and Yoshua Bengio. Nice: Non-linear independent components estimation. 2014.

[13] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Learning a deep convolutional network for image super-resolution. In *ECCV*, 2014.

[14] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Image super-resolution using deep convolutional networks. *TPAMI*, 38(2):295–307, 2016.

[15] Egor Ershov, Alex Savchik, Denis Shepelev, Nikola Banic, Michael S Brown, Radu Timofte, et al. NTIRE 2022 challenge on night photography rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2022.

[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.

[17] William T Freeman, Thouis R Jones, and Egon C Pasztor. Example-based super-resolution. *IEEE Computer graphics and Applications*, 2002.

[18] Jinjin Gu, Haoming Cai, Chao Dong, Jimmy Ren, Radu Timofte, et al. NTIRE 2022 challenge on perceptual image quality assessment. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2022.

[19] Shuhang Gu, Martin Danelljan, Radu Timofte, et al. Aim 2019 challenge on image extreme super-resolution: Methods and results. In *ICCV Workshops*, 2019.

[20] Muhammad Haris, Gregory Shakhnarovich, and Norimichi Ukita. Deep back-projection networks for super-resolution. In *CVPR*, 2018.

[21] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems*, 33:6840–6851, 2020.

[22] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models, 2020.

[23] Jia-Bin Huang, Abhishek Singh, and Narendra Ahuja. Single image super-resolution from transformed self-exemplars. In *CVPR*, 2015.

[24] Yiwen Huang and Ming Qin. Densely connected high order residual network for single frame image super resolution. *arXiv preprint arXiv:1804.05902*, 2018.

[25] Michal Irani and Shmuel Peleg. Improving resolution by image registration. *CVGIP*, 1991.

[26] Younghyun Jo, Sejong Yang, and Seon Joo Kim. Srflow-da: Super-resolution using normalizing flow with deep convolutional block. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 364–372, 2021.

[27] Hyeongju Kim, Hyeonseung Lee, Woo Hyun Kang, Joun Yeop Lee, and Nam Soo Kim. Softflow: Probabilistic framework for normalizing flow on manifolds, 2020.

[28] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee. Accurate image super-resolution using very deep convolutional networks. In *CVPR*, 2016.

[29] Younggeun Kim and Donghee Son. Noise conditional flow model for learning the super-resolution. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2021.

[30] Durk P Kingma and Prafulla Dhariwal. Glow: Generative flow with invertible 1x1 convolutions. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc., 2018.

[31] 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.

[32] 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.

[33] Ke Li, Shichong Peng, Tianhao Zhang, and Jitendra Malik. Multimodal image synthesis with conditional implicit maximum likelihood estimation. *International Journal of Computer Vision*, 128(10):2607–2628, 2020.

[34] Yawei Li, Kai Zhang, Radu Timofte, Luc Van Gool, et al. NTIRE 2022 challenge on efficient super-resolution: Methods and results. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2022.

[35] Jingyun Liang, Andreas Lugmayr, Kai Zhang, Martin Danelljan, Luc Van Gool, and Radu Timofte. Hierarchical conditional flow: A unified framework for image super-resolution and image rescaling. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4076–4085, 2021.

[36] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. Enhanced deep residual networks for single image super-resolution. *CVPR*, 2017.

[37] Zhi-Song Liu, Wan-Chi Siu, and Li-Wen Wang. Variational autoencoder for reference based image super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 516–525, 2021.

[38] Andreas Lugmayr, Martin Danelljan, Andres Romero, Fisher Yu, Radu Timofte, and Luc Van Gool. Repaint: Inpainting using denoising diffusion probabilistic models. *arXiv preprint arXiv:2201.09865*, 2022.

[39] Andreas Lugmayr, Martin Danelljan, Radu Timofte, et al. NTIRE 2021 learning the super-resolution space challenge. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2021.

[40] Andreas Lugmayr, Martin Danelljan, Radu Timofte, et al. NTIRE 2022 challenge on learning the super-resolution space. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2022.

[41] Andreas Lugmayr, Martin Danelljan, Luc Van Gool, and Radu Timofte. Srflow: Learning the super-resolution space with normalizing flow. In *ECCV*, pages 715–732. Springer, 2020.

[42] Andreas Lugmayr, Martin Danelljan, Fisher Yu, Luc Van Gool, and Radu Timofte. Normalizing flow as a flexible fidelity objective for photo-realistic super-resolution. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 1756–1765, 2022.

[43] Chenlin Meng, Yutong He, Yang Song, Jiaming Song, Jiajun Wu, Jun-Yan Zhu, and Stefano Ermon. Sdedit: Guided image synthesis and editing with stochastic differential equations. *arXiv preprint arXiv:2108.01073*, 2021.

[44] Shakir Mohamed and Balaji Lakshminarayanan. Learning in implicit generative models. *arXiv preprint arXiv:1610.03483*, 2016.

[45] Eliya Nachmani, Robin San Roman, and Lior Wolf. Denoising diffusion gamma models. *arXiv preprint arXiv:2110.05948*, 2021.

[46] Alex Nichol and Prafulla Dhariwal. Improved denoising diffusion probabilistic models. *arXiv preprint arXiv:2102.09672*, 2021.

[47] Alex Nichol, Prafulla Dhariwal, Aditya Ramesh, Pranav Shyam, Pamela Mishkin, Bob McGrew, Ilya Sutskever, and Mark Chen. Glide: Towards photorealistic image generation and editing with text-guided diffusion models. *arXiv preprint arXiv:2112.10741*, 2021.

[48] Sung Cheol Park, Min Kyu Park, and Moon Gi Kang. Super-resolution image reconstruction: a technical overview. *IEEE signal processing magazine*, 2003.

[49] Eduardo Pérez-Pellitero, Sibi Catley-Chandar, Richard Shaw, Ales Leonardis, Radu Timofte, et al. NTIRE 2022 challenge on high dynamic range imaging: Methods and results. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2022.

[50] Andres Romero, Angela Castillo, Jose M Abril-Nova, Radu Timofte, et al. NTIRE 2022 image inpainting challenge: Report. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2022.

[51] Chitwan Saharia, William Chan, Huiwen Chang, Chris A Lee, Jonathan Ho, Tim Salimans, David J Fleet, and Mohammad Norouzi. Palette: Image-to-image diffusion models. *arXiv preprint arXiv:2111.05826*, 2021.

[52] Jascha Sohl-Dickstein, Eric A. Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics, 2015.

[53] Libin Sun and James Hays. Super-resolution from internet-scale scene matching. In *ICCP*, 2012.

[54] 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.

[55] 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.

[56] Radu Timofte, Vincent De Smet, and Luc Van Gool. Anchored neighborhood regression for fast example-based super-resolution. In *ICCV*, pages 1920–1927, 2013.

[57] Longguang Wang, Yulan Guo, Yingqian Wang, Juncheng Li, Shuhang Gu, Radu Timofte, et al. NTIRE 2022 challenge on stereo image super-resolution: Methods and results. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2022.

[58] Xintao Wang, Ke Yu, Shixiang Wu, Jinjin Gu, Yihao Liu, Chao Dong, Chen Change Loy, Yu Qiao, and Xiaoou Tang. Esrgan: Enhanced super-resolution generative adversarial networks. *ECCV*, 2018.

[59] Valentin Wolf, Andreas Lugmayr, Martin Danelljan, Luc Van Gool, and Radu Timofte. Deflow: Learning complex image degradations from unpaired data with conditional flows. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 94–103, 2021.

[60] Chih-Yuan Yang and Ming-Hsuan Yang. Fast direct super-resolution by simple functions. In *ICCV*, pages 561–568, 2013.

[61] Jianchao Yang, John Wright, Thomas S. Huang, and Yi Ma. Image super-resolution as sparse representation of raw image patches. In *CVPR*, 2008.

[62] 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.

[63] Ren Yang, Radu Timofte, et al. NTIRE 2022 challenge on super-resolution and quality enhancement of compressed video: Dataset, methods and results. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2022.

[64] Xin Yu and Fatih Porikli. Ultra-resolving face images by discriminative generative networks. In *ECCV*, pages 318–333, 2016.

[65] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. *CVPR*, 2018.

[66] Hangqi Zhou, Chao Huang, Shangqi Gao, and Xiahai Zhuang. Vspsr: Explorable super-resolution via variational sparse representation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 373–381, 2021.