The Thirty-Eighth AAAI Conference on Artificial Intelligence (AAAI-24)

Uncertainty-Aware GAN for Single Image Super Resolution

Chenxi Ma
Fudan University, School of Computer Science and Technology, Shanghai, China
cxma@fudan.edu.cn

Abstract

Generative adversarial network (GAN) has become a popular tool in the perception-oriented single image super-resolution (SISR) for its excellent capability to hallucinate details. However, the performance of most GAN-based SISR methods is impeded due to the limited discriminative ability of their discriminators. In specific, these discriminators only focus on the global image reconstruction quality and ignore the more fine-grained reconstruction quality for constraining the generator, as they predict the overall realness of an image instead of the pixel-level realness. Here, we first introduce the uncertainty into the GAN and propose an **U**ncertainty-aware **GAN** (**UGAN**) to regularize SISR solutions, where the challenging pixels with large reconstruction uncertainty and importance (*e.g.*, texture and edge) are prioritized for optimization. The uncertainty-aware adversarial training strategy enables the discriminator to capture the pixel-level SR uncertainty, which constrains the generator to focus on image regions with high reconstruction difficulty, meanwhile, it improves the interpretability of the SR. To balance the weights of multiple training losses, we introduce an uncertainty-aware loss weighting strategy to learn the optimal loss weights adaptively. Extensive experiments demonstrate the effectiveness of our approach in extracting the SR uncertainty and the superiority of the UGAN over the state-of-the-art in terms of reconstruction accuracy and perceptual quality.

Introduction

Perception-oriented single image super-resolution (SISR), aiming to generate a high-resolution (HR) image with better perceptual quality from its low-resolution (LR) version, has gathered increasing attention. Thanks to the strong generation ability of the generative adversarial network (GAN), GAN-based SISR methods (SRGAN (Ledig et al. 2017), ESRGAN (Wang et al. 2018), RankSRGAN (Zhang et al. 2019), SPSR (Ma et al. 2020), Beby-GAN (Li et al. 2022), LDL (Liang, Zeng, and Zhang 2022), SROOE (Park, Moon, and Cho 2023)) can recover photo-realistic images.

However, the limited ability of the discriminator, which only predicts a compressed representation of the overall image quality, becomes a crucial obstacle to most of the current GAN-based SISR methods. Firstly, the discriminator treats all pixels in an image equally and fails to provide powerful feedback for the generator, resulting in unpleasant visual results (Figure 1). In addition, similar to most deep networks, these GAN models lack interpretability about the SR quality and reconstruction confidence of different pixels or regions. Therefore, GAN-based SISR methods call for a strong discriminator for further improvement.

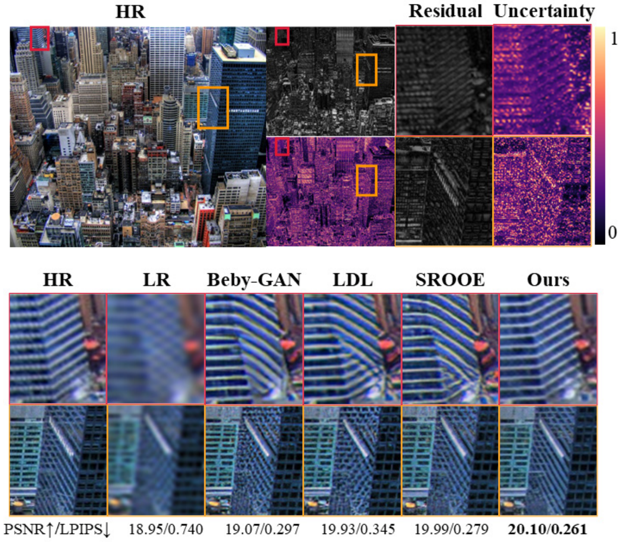

Figure 1: Comparison with state-of-the-art GAN-based SISR methods with the RRDBNet (Wang et al. 2018) backbone, including Beby-GAN (Li et al. 2022), LDL (Liang, Zeng, and Zhang 2022), and SROOE (Park, Moon, and Cho 2023). Our discriminator can predict SR uncertainty by indicating pixels with unreliable reconstruction performance.

In this paper, we propose an **U**ncertainty-aware **GAN** (**UGAN**) by learning the regionally different reconstruction uncertainty to regularize SISR solutions of the GAN. As (Qian et al. 2021) mentioned, pixels with large uncertainty carry visually important information and are hard to reconstruct, such as textures and edges. To focus on these challenging situations, we introduce the uncertainty into the adversarial training of the UGAN, where the discriminator can predict the SR uncertainty and enables the generator to pay more attention to image regions with high uncertainty

---

and SR difficulty by providing a more effective and fine-grained constraint. Our discriminator has another merit, that is, serving as an image quality assessment indicator for SR, as the uncertainty can also reflect the pixel-level SR quality. As shown in Figure 1, the UGAN enables better SR and stronger interpretability. Furthermore, how to balance weights between multiple losses in training is a key point of SR for a long time. We address this issue by introducing an uncertainty-aware loss weighting strategy to learn the optimal loss weights adaptively.

Our main contributions are summarized as follows.

- We introduce the uncertainty into the perception-oriented SR and propose an uncertainty-aware GAN (UGAN), resulting in better SR and stronger interpretability.
- We develop an uncertainty-aware adversarial training strategy, which enables the discriminator to learn the pixel-level uncertainty of the SR results and allows the generator to focus on image regions with higher reconstruction uncertainty.
- We propose an uncertainty-aware loss weighting strategy to balance contributions of different losses, which can avoid the troublesome effort of parameter tuning.

Related Work

Single Image Super Resolution

Benefiting from the development of deep learning, a huge number of convolutional neural network (CNN) based SISR approaches have been proposed and have achieved significant progress. The fidelity-oriented SISR methods (Kim, Lee, and Lee 2016; Lim et al. 2017; Zhang et al. 2018b; Wang et al. 2021a; Xia et al. 2022; Chen 2023) learn the LR-to-HR mapping by designing deep and complicated network architectures to improve the objective image quality (peak-to-noise ratio, PSNR).

Perception-Oriented SISR. To purchase better perceptual quality, SRGAN (Ledig et al. 2017) first introduces the GAN (Goodfellow et al. 2014) into SISR. ESRGAN (Wang et al. 2018) and Real-ESRGAN (Wang et al. 2021b) employ the relativistic discriminator to predict the relative realness of the SR image, compared to the ground truth (GT). SPSR (Ma et al. 2020) proposes a gradient discriminator to utilize the structural information of the gradient map. Liang *et al.* (Liang, Zeng, and Zhang 2022) proposed a locally discriminative learning (LDL) strategy to re-weigh the $\mathcal{L}_1$ loss through an artefact map. Since the traditional GAN or the relativistic average GAN (RaGAN) (Jolicoeur-Martineau 2019) are originally designed for the image generation task and are not perfectly suitable for SR, the above methods adopt the image-level discriminator to measure the overall image realness and rarely focus on the SR uncertainty. To reduce the GAN artefacts, Beby-GAN (Li et al. 2022) focuses on the quality of significant image regions by taking the foreground as input of the discriminator. However, Beby-GAN cannot provide per-pixel constraint for the generator, as its discriminator only outputs a single score to represent the overall image quality. SROOE (Park, Moon, and Cho 2023) employs objective trajectory learning to learn a target objective map, which controls each image patch to reflect the target SR performance obtained by the optimal combination of training losses (pixel-wise reconstruction loss, adversarial loss, and perceptual loss). To enhance the robustness of the discriminator, (Jo, Yang, and Kim 2020) and Real-ESRGAN adopt a U-net discriminator to predict a realness map of the input. Since the U-net discriminator takes a full-zero or full-one map as a label, its output still denotes the overall image quality instead of the local image quality.

Uncertainty in SR

Recently, uncertainty learning has been considered in SR. Kar *et al.* (Kar and Biswas 2021) proposed a Bayesian approach to estimate the uncertainty of SR results and defend the adversarial attacks. GRAM (Lee and Chung. 2019) decreases the loss attenuation of large variance pixels for better SR. Ning *et al.* (Qian et al. 2021) introduced the variance estimation to learn the SR uncertainty and proposed an uncertainty-driven loss (UDL) to improve the MSE loss. In the real-world SR field, (Z. Fang and Shi 2022) first introduces uncertainty learning to improve the robustness of the kernel estimation by learning the variance in the latent space of blur kernels. USR-DU (Ning et al. 2022) learns the degradation uncertainty of generated LR images with a Kullback-Leibler (KL) divergence-based loss and samples LR images from the uncertainty to construct LR-HR image pairs.

Different from the above SR methods, we explore the uncertainty in the GAN and enable the discriminator to learn the uncertainty of the SISR results and make it more interpretable for optimizing the generator.

Method

In this section, we propose an uncertainty-aware GAN (UGAN) for SR. First, we describe the network architecture and the overall training procedure of the proposed UGAN. Then, we discuss the uncertainty-aware adversarial training process in detail. Ultimately, we also introduce an uncertainty-aware loss weighting strategy to reduce the training complexity.

Overview

Network architecture. By following most GAN-based SR methods ((Wang et al. 2018; Ma et al. 2020; Li et al. 2022; Liang, Zeng, and Zhang 2022)), the proposed UGAN is composed of a generator and a discriminator. During inference, only a feed-forward calculation of the generator is required to reconstruct the HR image. In general, the epistemic uncertainty of the generator can be represented by indicating unreliable pixels in the reconstructed image, and the lower uncertainty always leads to better SR. To denote the per-pixel uncertainty of the SR results, the discriminator produces an image reconstruction confidence map and provides fine-grained feedback for constraining the generator. Thereby, the generator can recover realistic image details with less uncertainty through an uncertainty-aware adversarial loss, which will be described in detail later.

The generator can adopt any differentiable SR network and we choose three representative SR backbones (SRResNet (Ledig et al. 2017), RRDBNet (Wang et al. 2018),

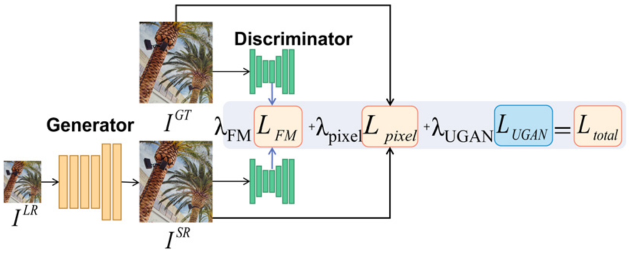

Figure 2: The total training loss of the UGAN, including the pixel loss ($\mathcal{L}_{pixel}$), the feature similarity loss ($\mathcal{L}_{FM}$), and the adversarial loss ($\mathcal{L}_{UGAN}$).

SwinIR (Liang et al. 2021)) in the experiment. As for the discriminator, we adopt the U-net structure, where the encoder consists of six 3×3 convolutional layers with stride 2 and the decoder has six upsampling layers followed by six 3×3 convolutional layers with stride 1. The feature number of the middle convolutional layers is 64.

Note that, the proposed discriminator can serve as a no-reference (NR) pixel-level image quality assessment indicator for SR (SR-QA), in addition to contributing an effective adversarial loss. It is generally believed that the reconstruction confidence of a high-quality SR result is high, corresponding to the low uncertainty. Therefore, the simple average of the discriminator's output can represent the overall image quality, the lower the better. In this work, we will briefly verify this ability of our discriminator in experiments, providing several insights for further applying the proposed approach to other IR tasks. A more appropriate approach for calculating the quality score from the output of the discriminator may further improve the SR-QA accuracy and will be studied in our future work.

Total training loss. As shown in Figure 2, the total loss function ($\mathcal{L}_{total}$) of the generator in our UGAN contains the pixel loss ($\mathcal{L}_{pixel}$), the feature matching loss (Jo, Yang, and Kim 2020) ($\mathcal{L}_{FM}$), and the uncertainty-aware adversarial loss ($\mathcal{L}_{UGAN}$), which will be introduced in detail later.

$$\mathcal{L}_{total} = \lambda_{UGAN}\mathcal{L}_{UGAN} + \lambda_{pixel}\mathcal{L}_{pixel} + \lambda_{FM}\mathcal{L}_{FM}, \tag{1}$$

where $\lambda_{UGAN}, \lambda_{pixel}, \lambda_{FM}$, denote the weighting parameters to balance the effect of different loss functions.

Suppose (x, y) denotes a pair of LR image and HR ground truth (GT) image, $f_G(\cdot)$ and $f_D(\cdot)$ are operations of the generator and the discriminator. We adopt the L2 norm ($L2(\cdot)$) for the $\mathcal{L}_{pixel}$ to calculate the pixel-wise similarity between the $f_G(x)$ and y.

$$\mathcal{L}_{pixel} = ||f_G(x) - y||_2, \tag{2}$$

By following the work (Jo, Yang, and Kim 2020), the perceptual loss $\mathcal{L}_{FM}$ constrains the feature similarity between $f_G(x)$ and y based on the pre-trained classification network $\phi(\cdot)$.

$$\mathcal{L}_{FM} = ||\phi(f_G(x)) - \phi(y)||_2. \tag{3}$$

Uncertainty-Aware Adversarial Training

To make the discriminator learn the uncertainty of the reconstructed image and make the generator focus on pixels with high uncertainty, we propose an uncertainty-aware adversarial training strategy to optimize the discriminator and the generator based on the regionally different reconstruction uncertainty. Specifically, to calculate the uncertainty-aware adversarial loss ($\mathcal{L}_{UGAN}$), we generate the training data through uncertainty-aware data augmentation. In addition to the $\mathcal{L}_{UGAN}$, other training losses take the unprocessed HR and SR images $(y, f_G(x))$ as input.

Uncertainty-aware data augmentation. As shown in Figure 3, the training data ($I = [I^D, I^G], M = [M^D, M^G]$) for the adversarial training is generated by sampling pixels from the SR result (I^{SR}) and the GT (I^{GT}) based on two probability distributions, where I denotes the input image of the discriminator and M denotes the label of the per-pixel image realness. The superscripts (D, G) of I and M indicate the training data for the discriminator or the generator. The zero map is used as the training label M^G for the generator.

The visually important image regions, such as textures and edges, have large uncertainty and are hard to reconstruct (Qian et al. 2021). Therefore, the large reconstruction errors (residual between the SR result and the GT) can denote unreliable reconstruction performance and large uncertainty. We first obtain the residual ($I^{Residual}$) between the I^{GT} and the I^{SR} and normalize the $I^{Residual}$.

Next, we generate probability maps ($P = [P^{Proportion}, P^{Gauss}]$) by applying the proportional distribution and the Gaussian distribution on the $I^{Residual}$, as described in the following formulas:

$$\begin{aligned} P^{Proportion}_{i,j} &= f_{prop}(R_{i,j}) = f_{identity}(\mathcal{N}orm(R_{i,j})), \\ P^{Gauss}_{i,j} &= f_{Gauss}(R_{i,j}, 0.5, \tfrac{1}{6}^2), \end{aligned} \tag{4}$$

where $R_{i,j}$ is the pixel value (i, j) of the $I^{Residual}$, $f_{identity}(\cdot)$ is the identity mapping, $\mathcal{N}orm(\cdot)$ is the operation of normalization, $f_{prop}(\cdot)$ is the proportional probability density function, and $f_{Gauss}(\cdot)$ is the Gaussian probability density function with the mathematic expectation $\mu = 0.5$ and the variance $\sigma^2 = \frac{1}{6}^2$.

Based on the probability map P, we sample pixels in the I^{GT} and the I^{SR} T times to synthesize training images ($I^t,\ _{t=1,\ldots,T}$).

$$\begin{aligned} I^t_{i,j} &= \begin{cases} I^{SR}_{i,j} & p \sim P_{i,j} \\ I^{GT}_{i,j} & p \sim 1 - P_{i,j} \end{cases}, \\ M^t_{i,j} &= \begin{cases} 1 & p \sim P_{i,j} \\ 0 & p \sim 1 - P_{i,j} \end{cases}, \end{aligned} \tag{5}$$

where $p \sim P_{i,j}$ denotes the probability that $I_{i,j}$ is sampled from the pixel (i, j) in I^{SR}. Such a sampling process can be regarded as combining I^{GT} and I^{SR} according to the sampling masks M^t (Eq.(6)), where $M_{i,j} = 1$ indicate that the pixel $I_{i,j}$ belongs to the reconstructed image. Finally, the synthesized images and masks serve as the training data for the adversarial training of the UGAN.

$$I^t = I^{SR} * M^t + I^{GT} * (1 - M^t). \tag{6}$$

Uncertainty-aware adversarial loss. Given N pairs of training samples ($I^{D^n}, M^{D^n}, I^{G^n}, n = 1, \ldots, N$), the

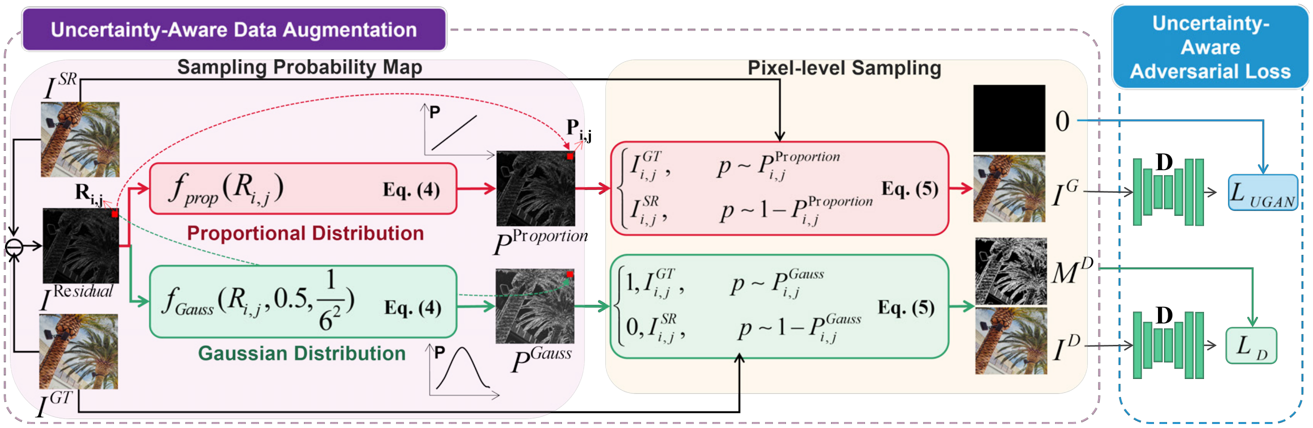

Figure 3: Overview of the proposed uncertainty-aware GAN (UGAN). Pipeline of the uncertainty-aware data augmentation, used to generate the training data (I^G, I^D, M^D) for calculating the $\mathcal{L}_{UGAN}$ and $\mathcal{L}_D$. Based on the residual ($I^{Residual}$) between the SR image (I^{SR}) and the GT (I^{GT}), the pixel-level sampling probability maps ($P^{Proportion}$, P^{Gauss}) are generated through the proportional and Gaussian probability distribution functions in Eq.(4). Then, we sample pixels of I^{SR} and I^{GT} based on $P^{Proportion}$ and P^{Gauss} to generate fused images (I^G, I^D) and a mask (M^D) through the operation in Eq.(5). The adversarial losses ($\mathcal{L}_{UGAN}, \mathcal{L}_D$) for the generator and the discriminator calculate the accuracy of the output ($f_D(I^G), f_D(I^D)$) of the discriminator by taking a 0-matrix and the mask (M^D) as reference.

uncertainty-aware adversarial losses ($\mathcal{L}_D, \mathcal{L}_{UGAN}$) for the discriminator and the generator are defined as:

$$\begin{aligned} \mathcal{L}_D &= \textstyle\sum_{i,j,n}(|f_D(I^{D^n})_{i,j} - M^{D^n}_{i,j}|), \\ \mathcal{L}_{UGAN} &= \textstyle\sum_{i,j,n}(|f_D(I^{G^n})_{i,j} - 0|), \end{aligned} \tag{7}$$

where $\mathcal{L}_D$ takes the sampling mask M^D as label to calculate the pixel-level constraint and $\mathcal{L}_{UGAN}$ takes a zero map as label to ensure the consistency of pixels in I^G, which enables the pixel-level similarity between output (I^{SR}) of the generator and the GT (I^{GT}).

Discussions. We adopt different probability distributions to synthesize the training data ([I^D, M^D], [I^G]) for the discriminator and the generator based on the following considerations: Firstly, we hope the generator focuses on pixels with large reconstruction uncertainty, that are hard to reconstruct, thereby we adopt the proportional distribution to generate I^G. Thus, pixels with larger residuals (reconstruction error) are more likely to be sampled to generate the training data for L_{UGAN}. Secondly, to make the discriminator have strong discrimination ability and be sensitive to extensive image regions with different degrees of reconstruction error, we sample pixels based on the probability distribution (Gaussian) of the pixel number. Since the pixels with larger or smaller residual values located in the range of [0, 1], gradually decrease, we set $\mu, \sigma = 0.5, \frac{1}{6}$ (confidence interval $[\mu - 3\sigma, \mu + 3\sigma] = [0, 1]$) to make the discriminator be trained with more reconstructed pixels, the residual values of which are in the middle range.

Compared with existing GAN-based SR methods, there are mainly the following differences in our UGAN. Firstly, the previous discriminators often output a single value or a 2-D map to represent the global image quality, our discriminator measures the pixel-wise SR uncertainty to provide accurate and fine-grained feedback for the generator. Secondly, instead of directly inputting one SR (GT) image to the discriminator, the uncertainty-aware data augmentation fuses pixels in SR and HR images to generate a new input for the discriminator by considering the pixel-level SR accuracy. In this way, the discriminator can learn to be a no-reference image quality assessment model, that focuses on the per-pixel SR performance, and the generator can reasonably focus on pixels that are difficult to reconstruct.

Uncertainty-Aware Loss Weighting

Choosing appropriate weights λ in Eq.(1) for different losses is significant, as the loss weight greatly influences the final model. However, it is difficult to balance multiple losses in training and most of the existing SR methods set the loss weights by experience, which increases the training difficulty and may be harmful to the final results.

Inspired by the multi-task learning work (Kendall, Gal, and Cipolla 2018), which applies the task's uncertainty to weigh different losses, we also explore the possibility of learning the optimal weights to re-weigh the power of different losses for the SR task based on the loss uncertainty. We consider that the loss uncertainty reflects the confidence of the loss and depends on the calculation manners and evaluation metrics of the loss.

Assuming the output of a generator follows the Gaussian distribution. We define SR result ($f_G(x)$) as the mean of the Gaussian likelihood with an observation noise scalar σ. In maximum likelihood inference, we maximize the log-likelihood of the generator.

$$\begin{aligned} &p(y|f_G(x)) = \mathcal{N}(f_G(x), \sigma^2), \\ &\log p(y|f_G(x)) \propto -\tfrac{1}{2\sigma^2}\mathcal{L}(y, f_G(x)) - \log \sigma, \end{aligned} \tag{8}$$

where $\mathcal{L}(\cdot)$ denotes the function used to calculate the distance between y and $f_G(x)$ and can be replaced with dif-

Backbone	Methods	Set14	BSD100	Urban100	DIV2K100
Fidelity-Oriented	Bicubic	25.762/0.7029/0.4379	25.961/0.6683/0.5219	23.121/0.6583/0.4725	27.352/0.7697/0.4192
	UDL	28.003/0.7922/0.2936	26.289/0.7242/0.3884	25.520/0.7814/0.2537	29.764/0.9014/0.1158
	ENLCN	28.940/0.7892/0.3357	27.820/0.7452/0.4315	27.120/0.8141/0.3484	30.539/0.8523/0.2631
	HAT	29.152/0.7958/0.2510	27.935/0.7505/0.3492	27.873/0.8346/0.1773	32.350/0.9283/0.0897
SRResNet	SRGAN	26.839/0.7252/0.1327	24.378/0.6195/0.2036	24.338/0.7235/0.1439	26.766/0.7570/0.1619
	RankSRGAN	26.456/0.7055/0.1374	25.511/0.6530/0.1732	24.488/0.7319/0.1380	28.093/0.7854/0.1208
	UGAN	27.381/0.7405/0.0756	26.479/0.6944/0.0886	25.184/0.7590/0.0967	29.093/0.8026/0.0538
RRDBNet	ESRGAN	26.064/0.6971/0.1341	25.389/0.6480/0.1615	24.367/0.7324/0.1231	28.169/0.7898/0.1154
	SPSR	26.345/0.7115/0.1323	25.533/0.6581/0.1617	24.779/0.7465/0.1192	28.177/0.7861/0.1099
	BebyGAN	26.960/0.7282/0.1268	25.810/0.6781/0.1613	23.357/0.7041/0.1287	28.620/0.7900/0.1022
	LDL	27.228/0.7358/0.1132	25.970/0.6794/0.1563	25.459/0.7661/0.1084	28.951/0.7951/0.0504
	SROOE	26.997/0.7496/0.1150	24.870/0.6869/0.1500	24.330/0.7707/0.1065	27.690/0.7932/0.0957
	UGAN	**27.935/0.7583**/0.0732	**26.903/0.7144**/0.0859	25.950/0.7858/0.0858	29.381/0.7965/**0.0464**
SwinIR	LDL	27.526/0.7478/0.1091	26.305/0.6929/0.1623	26.231/0.7918/0.1021	29.117/0.8011/0.0485
	UGAN	27.674/0.7477/**0.0675**	26.763/0.7090/**0.0805**	**26.390/0.7956/0.0779**	**29.514/0.8127**/0.0476

Table 1: Quantitative comparison with state-of-the-art fidelity-oriented (UDL (Qian et al. 2021), ENLCN (Xia et al. 2022), HAT (Chen 2023)) and GAN-based SISR approaches (SRGAN (Ledig et al. 2017), ESRGAN (Wang et al. 2018), RankSRGAN (Zhang et al. 2019), SPSR (Ma et al. 2020), Beby-GAN (Li et al. 2022), LDL (Liang, Zeng, and Zhang 2022), SROOE (Park, Moon, and Cho 2023)). Three groups of comparisons are made based on the employed backbones: SRResNet (Ledig et al. 2017), RRDBNet (Wang et al. 2018), and SwinIR (Liang et al. 2021). The average PSNR↑/SSIM↑/LPIPS↓ of SR results on benchmarks, ↑ denotes the higher, the better. The best results of GAN-based approaches are highlighted in bold.

ferent loss functions in Eq.(2), Eq.(3), Eq.(7) to obtain the minimization objective $\mathcal{L}_{total}$ for the SR model.

$$\begin{aligned}\mathcal{L}_{total} &= -\log[p(y|G(x))],\\ &\propto \frac{1}{2\sigma^2}\mathcal{L}(y, f_G(x)) + \log\sigma,\\ &= \frac{1}{2\sigma_1^2}\mathcal{L}_{pixel}(y, f_G(x)) + \frac{1}{2\sigma_2^2}\mathcal{L}_{FM}(y, f_G(x))\\ &\quad + \frac{1}{2\sigma_3^2}\mathcal{L}_{UGAN}(f_G(x)) + \log\sigma_1\sigma_2\sigma_3.\end{aligned} \tag{9}$$

After training with the $\mathcal{L}_{total}$, the optimal σ can be obtained to calculate the weight ($\lambda = \frac{1}{2\sigma^2}$) for each loss.

Experiments

This section mainly introduces the experimental settings and reports the results of the UGAN. We compare the UGAN with state-of-the-arts and conduct the ablation study to verify the effect of different designs. We also explore the potential of the discriminator on the SISR quality assessment (SR-QA).

Database and Protocols

By following the GAN-based SR methods (Wang et al. 2018; Zhang et al. 2019; Liang, Zeng, and Zhang 2022; Li et al. 2022; Park, Moon, and Cho 2023), we train our models on the DF2K datasets, which is composed of DIV2K (Timofte et al. 2017b) (800 images) and Flickr2K (Timofte et al. 2017a) (2650 images). Training images are augmented by randomly down-scaling, flipping, and rotating. We obtain the LR image by bicubic interpolation and consider the scaling factor 4 in experiments. Each mini-batch contains 32 HR image patches of size 120×120.

For evaluation, we adopt Set5 (Bevilacqua et al. 2012), Set14 (Zeyde, Elad, and Protter 2010), BSD100 (Martin et al. 2001), Urban100 (Huang, Singh, and Ahuja 2015), Manga109 (Yusuke Matsui and Aizawa 2017), and DIV2K100 (Timofte et al. 2017b) datasets and common metrics, including Peak Signal-to-Noise Ratio (PSNR), Structure Similarity (SSIM) (Zhou Wang et al. 2004), and Learned Perceptual Image Patch Similarity (LPIPS) (Zhang et al. 2018a). The PSNR and SSIM are calculated on the brightness channel and the LPIPS is calculated on the RGB channels. The higher PSNR/SSIM and the lower LPIPS indicate better results. All models are based on the TensorFlow (Abadi et al. 2015) implementation and optimized by Adam (P. and Ba 2015) with $\beta 1 = 0.9$, $\beta 2 = 0.999$. The learning rate is 1e-4. All experiments are conducted on a machine with a Nvidia GTX2080Ti GPU (128G RAM).

Comparison with Prior Art

SISR. The proposed UGAN is compared with state-of-the-art GAN-based (SRGAN (Ledig et al. 2017), ESRGAN (Wang et al. 2018), RankSRGAN (Zhang et al. 2019), SPSR (Ma et al. 2020), Beby-GAN (Li et al. 2022), LDL (Liang, Zeng, and Zhang 2022), SROOE (Park, Moon, and Cho 2023)) and fidelity-oriented (UDL (Qian et al. 2021), ENLCN (Xia et al. 2022), HAT (Chen 2023)) SISR methods on several benchmarks. The quantitative comparison is reported in Table 1. The fidelity-oriented SISR methods always lead to high PSNR/SSIM accompanied by poor perceptual performance. For a fair comparison, we adopt three groups of comparisons based on different backbones (SRResNet (Ledig et al. 2017), RRDBNet (Wang et al. 2018), SwinIR (Liang et al. 2021)). Among all competing GAN-based methods, the UGAN achieves the best PSNR/SSIM/LPIPS. In Figure 4 and Figure 5, the UGAN produces accurate details and clear structure, that are clos-

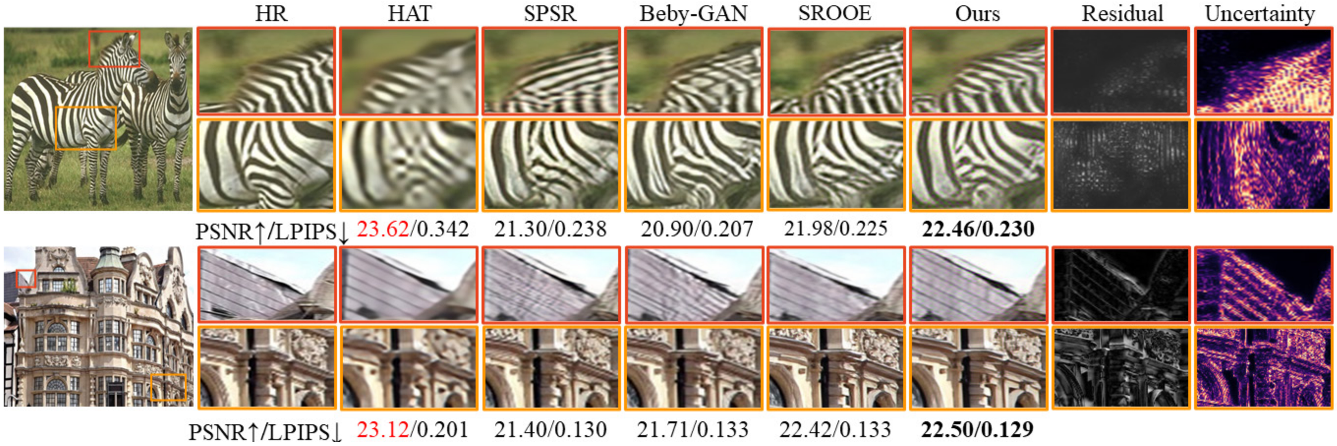

Figure 4: Visual comparison with state-of-the-art SISR approaches (HAT (Chen 2023), SPSR (Ma et al. 2020), Beby-GAN (Li et al. 2022), and SROOE (Park, Moon, and Cho 2023)). Better zoom in.

er to the GT, especially for some challenging SR situations (textures, edges) with large reconstruction uncertainty.

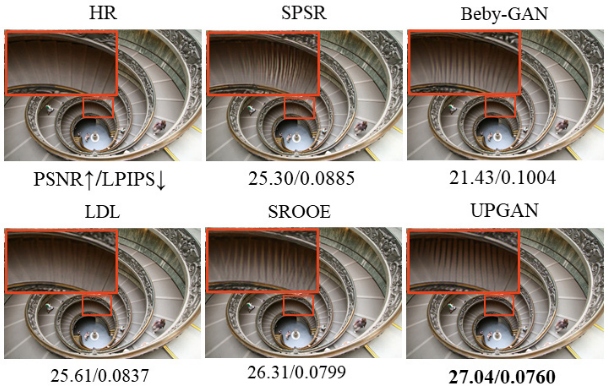

Figure 5: Visual comparison with GAN-based SR methods with the RRDBNet backbone, including SPSR, Beby-GAN, LDL, and SROOE.

Gaussian denoising and deblurring. We extend the U-GAN, GAN and LDL with RRDBNet backbone to the Gaussian colour image denoising and deblurring. Table 2 reports the average PSNR/SSIM/LPIPS on the BSD100 for noise levels σ=[15,25,35] and the average PSNR/SSIM/LPIPS on the BSD100 for three isotropic and three anisotropic Gaussian blur kernels, the visualized kernels are shown in the supplementary file.

Real-world SISR and denoising. We further apply the proposed UGAN to the real-world LR images, suffering from the unknown and complicated degradation, by retraining the RRDBNet+UGAN model on the same training data of the RealESRGAN (Wang et al. 2021c). Figure 6 compares the $\times4$ SR results. We also conduct real image denoising experiments by training the UGAN, GAN and LDL with RRDBNet backbone on the real-world datasets SIDD (Abdelhamed et al. 2019), and demonstrate the average PSNR/SSIM in Table 2.

σ	GAN	LDL	UGAN (ours)
15	34.19/0.0246	33.98/0.0247	**34.64/0.0222**
25	31.72/0.0412	31.68/0.0397	**32.07/0.0346**
35	30.09/0.0601	30.12/0.0567	**30.44/0.0476**
SIDD	35.73/0.907	35.85/0.919	**35.98/0.923**
iso-blur	32.33/**0.080**	31.70/0.086	**32.42**/0.082
aniso-blur	29.58/0.101	29.43/0.105	**29.68/0.096**

Table 2: PSNR/LPIPS results of Gaussian colour image denoising (σ=[15, 25, 35]), isotropic, and anisotropic Gaussian deblurring on BSD100. PSNR/SSIM results of real image denoising on SIDD.

Figure 6: The $\times4$ SR results on real-world LR samples.

Ablation Study

We experiment with different design options to analyze the performance of our key contributions. All models in the ablation study adopt the SRResNet backbone.

Training loss. As shown in Table 3, we first verify the effectiveness of different training losses in Eq.(1).

Loss	ULW	BSD100	Urban100
$\mathcal{L}_{pixel}$	×	**27.58/0.762**/0.3758	**26.09/0.786**/0.2255
$+\mathcal{L}_{FM}$	×	26.62/0.698/0.1243	24.95/0.748/0.1251
$+\mathcal{L}_{UGAN}$	×	26.48/0.694/**0.0886**	25.18/0.759/**0.0967**
$+\mathcal{L}_{UGAN}$	✓	26.54/0.688/0.1584	25.61/0.764/0.1243

Table 3: Ablation study on the training loss and the ULW.

Sampling Distribution		Pixel/Patch	PSNR/SSIM/LPIPS
$\mathcal{L}_{UGAN}$	$\mathcal{L}_D$		
P^{no}	P^{no}	Pixel	25.056/0.655/0.349
$P^{uniform}$	$P^{uniform}$	Patch	26.707/0.707/0.290
$P^{uniform}$	$P^{uniform}$	Pixel	26.182/0.701/0.360
$P^{Proportion}$	$P^{Proportion}$	Pixel	25.116/0.647/0.261
P^{Gauss}	P^{Gauss}	Pixel	26.086/0.681/0.318
$P^{Proportion}$	P^{exp}	Pixel	26.214/0.722/0.373
P^{exp}	P^{Gauss}	Pixel	26.525/0.703/0.332
$P^{Proportion}$	P^{Gauss}	Pixel	26.479/0.694/0.089

Table 4: Ablation study of the sampling distribution in the UDA for the generator and the discriminator.

Generators		BSD100	Urban100
SRResNet	GAN	24.38/0.620/0.204	24.34/0.724/0.144
	UGAN	26.48/0.694/0.089	25.18/0.759/0.097
RRDBNet	GAN	26.50/0.695/0.104	24.92/0.752/0.124
	UGAN	**26.90/0.714**/0.086	25.95/0.786/0.086
SwinIR	GAN	26.66/0.697/0.096	25.48/0.767/0.094
	UGAN	26.76/0.709/**0.081**	**26.39/0.796/0.078**

Table 5: The generalization ability of the GAN and our UGAN on different backbones.

Uncertainty-aware loss weighting (ULW). The loss weight often influences the final SR result, and it is difficult to set the best weight for balancing different losses. In Table 3, though the ULW does not lead to the best SR, it can reduce the training burden by adaptively learning the loss weight to avoid the effort of parameter tuning, accompanied by competitive SR performance.

Uncertainty-aware data augmentation (UDA) for the adversarial training. To verify the effect of the UDA, we design different data augmentation methods by adopting different sampling probability P in Eq.(5). In specific, the probability distributions include the identical probability distributions with $P^{no} = 0$ and $P^{uniform} = 0.5$, the exponential distribution $P^{exp} = e^{-10x}$, the proportional distribution $P^{Proportion}$, and the Gaussian distribution P^{Gauss}. Then, we combine different P to generate the training data (I, M in Eq.(7)) for calculating $\mathcal{L}_{UGAN}$ and $\mathcal{L}_D$ and adopt the total loss in Eq.(1) to train different GAN-based SR models with the same architecture. Table 4 shows the quantitative results of the above models on the BSD100.

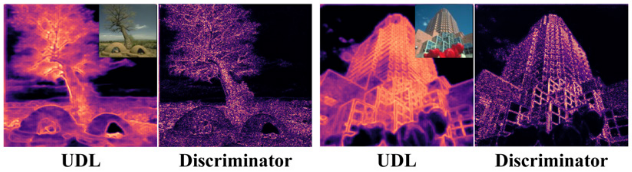

Figure 7: Visualization of the uncertainty, predicted by UDL (Qian et al. 2021) and our discriminator, for ×4 SR.

IQA	SROCC↑	KROCC↑	PLCC↑	RMSE↓
PSNR	0.575	0.403	0.592	0.215
SSIM	0.569	0.396	0.591	0.216
FSIM	0.700	0.507	0.721	0.185
LPIPS	0.706	0.510	0.740	0.180
SISRQA	0.984	0.904	0.988	0.041
ours	**0.787**	**0.583**	**0.819**	**0.153**

Table 6: Quantitative results on the SR-QA task.

Analysis of Discriminator

Backbone of the UGAN. We compare the standard GAN and the UGAN based on different backbones (SRResNet (Ledig et al. 2017), RRDBNet (Wang et al. 2018), SwinIR (Liang et al. 2021)) in Table 5. The UGAN can flexibly drive these backbones to achieve better SR performance.

Uncertainty estimation. Figure 7 compares the uncertainty output of our discriminator and the UDL (Qian et al. 2021). Our discriminator can accurately indicate pixels with large reconstruction uncertainty.

SR quality assessment (SR-QA). As we mentioned before, our discriminator has an extra merit, being regarded as an NR-image quality assessment (IQA) model for SR. Table 6 compares common IQA methods (PSNR, SSIM, FSIM (Zhang et al. 2011), LPIPS (Zhang et al. 2018a), SISRQA (Yan et al. 2019)) on the SR-QA dataset (Ma et al. 2017), and indicates the SROCC, KROCC, PLCC, and RMSE results. The CNN-based SISRQA model (Yan et al. 2019), trained on the SR-QA dataset, achieves the best results. Though our discriminator is not specifically designed for the IQA task nor trained on the SR-QA dataset, it achieves promising performance. A better way to calculate the quality score from the output of the discriminator may lead to a higher SR-QA accuracy and will be studied in our future work. In addition, our discriminator is more practical, as it does not require a clean HR image as a reference.

Future Work

Diffusion models (Sohl-Dickstein and Ganguli 2015; Ho and Abbeel 2020; Song and Kautz 2023; Song and Ermon 2020; Yang et al. 2023) have achieved unprecedented success across multiple fields and show great potential in image restoration. Therefore, how to explore the uncertainty of the diffusion model for better SR and higher interpretability may be a promising topic and will be further studied in future.

Conclusion

This paper introduced the uncertainty into the GAN-based image super-resolution and proposed a novel uncertainty-aware GAN (UGAN) for reconstructing well-textured images. Specifically, we designed the uncertainty-aware adversarial training strategy to constrain the UGAN based on the reconstruction uncertainty of the SR results. Therefore, our UGAN can predict the pixel-level reconstruction uncertainty, making it more interpretable, and meanwhile, it can focus on image regions with large uncertainty during training to obtain better SR. Extensive experiments demonstrate the superiority of our UGAN over the state-of-the-art.

References

Abadi, M.; Agarwal, A.; Barham, P.; Brevdo, E.; Chen, Z.; Citro, C.; Corrado, G. S.; Davis, A.; Dean, J.; Devin, M.; Ghemawat, S.; Goodfellow, I.; Harp, A.; Irving, G.; Isard, M.; Jia, Y.; Jozefowicz, R.; Kaiser, L.; Kudlur, M.; Levenberg, J.; Mané, D.; Monga, R.; Moore, S.; Murray, D.; Olah, C.; Schuster, M.; Shlens, J.; Steiner, B.; Sutskever, I.; Talwar, K.; Tucker, P.; Vanhoucke, V.; Vasudevan, V.; Viégas, F.; Vinyals, O.; Warden, P.; Wattenberg, M.; Wicke, M.; Yu, Y.; and Zheng, X. 2015. TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems.

Abdelhamed, A.; Timofte, R.; Brown, M. S.; et al. 2019. NTIRE 2019 Challenge on Real Image Denoising: Methods and Results. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*.

Bevilacqua, M.; Roumy, A.; Guillemot, C.; and Alberi-Morel, M. L. 2012. Low-complexity single-image super-resolution based on nonnegative neighbor embedding.

Chen, X. . Z. J. . Q. Y. . D. C., Xiangyu ; Wang. 2023. Activating More Pixels in Image Super-Resolution Transformer. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.

Goodfellow, I. J.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; and Bengio, Y. 2014. Generative Adversarial Networks. In *Advances in Neural Information Processing Systems (NeurIPS)*.

Ho, J. A., J.; and Abbeel, P. 2020. Denoising diffusion probabilistic models. In *Advances in Neural Information Processing Systems (NeurIPS)*.

Huang, J.-B.; Singh, A.; and Ahuja, N. 2015. Single image super-resolution from transformed self-exemplars. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 5197–5206.

Jo, Y.; Yang, S.; and Kim, S. J. 2020. Investigating Loss Functions for Extreme Super-Resolution. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 1705–1712.

Jolicoeur-Martineau, A. 2019. The relativistic discriminator: a key element missing from standard GAN. *The International Conference on Learning Representations (ICLR)*.

Kar, A.; and Biswas, P. K. 2021. Fast Bayesian Uncertainty Estimation and Reduction of Batch Normalized Single Image Super-Resolution Network. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.

Kendall, A.; Gal, Y.; and Cipolla, R. 2018. Multi-Task Learning Using Uncertainty to Weigh Losses for Scene Geometry and Semantics. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.

Kim, J.; Lee, J. K.; and Lee, K. M. 2016. Accurate Image Super-Resolution Using Very Deep Convolutional Networks. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 1646–1654.

Ledig, C.; Theis, L.; Huszr, F.; Caballero, J.; Cunningham, A.; Acosta, A.; Aitken, A.; Tejani, A.; Totz, J.; Wang, Z.; and Shi, W. 2017. Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 105–114.

Lee, C.; and Chung., K.-S. 2019. Gram: Gradient rescaling attention model for data uncertainty estimation in single image super resolution. In *IEEE International Conference On Machine Learning And Applications (ICMLA)*.

Li, W.; Zhou, K.; Qi, L.; Lu, L.; Jiang, N.; Lu, J.; and Jia, J. 2022. Best-Buddy GANs for Highly Detailed Image Super-Resolution. In *Association for the Advancement of Artificial Intelligence (AAAI)*.

Liang, J.; Cao, J.; Sun, G.; Zhang, K.; and Timofte, R. 2021. SwinIR: Image Restoration Using Swin Transformer. In *IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*.

Liang, J.; Zeng, H.; and Zhang, L. 2022. Details or Artifacts: A Locally Discriminative Learning Approach to Realistic Image Super-Resolution. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.

Lim, B.; Son, S.; Kim, H.; Nah, S.; and Lee, K. M. 2017. Enhanced deep residual networks for single image super-resolution. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, volume 1, 4.

Ma, C.; Rao, Y.; Cheng, Y.; Chen, C.; Lu, J.; and Zhou, J. 2020. Structure-Preserving Super Resolution with Gradient Guidance. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.

Ma, C.; Yang, C. Y.; Yang, X.; and Yang, M. H. 2017. Learning a No-Reference Quality Metric for Single-Image Super-Resolution. *Computer Vision & Image Understanding*, 158: 1–16.

Martin, D.; Fowlkes, C.; Tal, D.; and Malik, J. 2001. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *IEEE International Conference on Computer Vision (ICCV)*, volume 2, 416–423. IEEE.

Ning, Q.; Tang, J.; Wu, F.; Dong, W.; Li, X.; and Shi, G. 2022. Learning degradation Uncertainty for unsupervised real-world image super-resolution. In *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence (IJCAI)*.

P., K. D.; and Ba, J. L. 2015. Adam: A method for stochastic optimization. In *International Conference on Learning Representations (ICLR)*.

Park, S. H.; Moon, Y. S.; and Cho, N. I. 2023. Perception-Oriented Single Image Super-Resolution Using Optimal Objective Estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 1725–1735.

Qian, N.; Dong, W.; Li, X.; Wu, J.; and Shi, G. 2021. Uncertainty-Driven Loss for Single Image Super-Resolution. In *Advances in Neural Information Processing Systems (NeurIPS)*.

Sohl-Dickstein, W. E. M. N., J.; and Ganguli, S. 2015. Deep Unsupervised Learning Using Nonequilibrium Thermodynamics. In *International Conference on Machine Learning (ICML)*.

Song, V. A. M. M., J.; and Kautz, J. 2023. Pseudoinverse-guided diffusion models for inverse problems. In *International Conference on Learning Representations (ICLR)*.

Song, Y.; and Ermon, S. 2020. Improved Techniques for Training Score-Based Generative Models. In *Advances in Neural Information Processing Systems (NeurIPS)*.

Timofte, R.; Agustsson, E.; Gool, L. V.; Yang, M.-H.; Zhang, L.; Lim, B.; Son, S.; Kim, H.; Nah, S.; Lee, K. M.; Wang, X.; Tian, Y.; Yu, K.; Zhang, Y.; Wu, S.; Dong, C.; Lin, L.; Qiao, Y.; Loy, C. C.; Bae, W.; Yoo, J.; Han, Y.; Ye, J. C.; Choi, J.-S.; Kim, M.; Fan, Y.; Yu, J.; Han, W.; Liu, D.; Yu, H.; Wang, Z.; Shi, H.; Wang, X.; Huang, T. S.; Chen, Y.; Zhang, K.; Zuo, W.; Tang, Z.; Luo, L.; Li, S.; Fu, M.; Cao, L.; Heng, W.; Bui, G.; Le, T.; Duan, Y.; Tao, D.; Wang, R.; Lin, X.; Pang, J.; Xu, J.; Zhao, Y.; Xu, X.; Pan, J.; Sun, D.; Zhang, Y.; Song, X.; Dai, Y.; Qin, X.; Huynh, X.-P.; Guo, T.; Mousavi, H. S.; Vu, T. H.; Monga, V.; Cruz, C.; Egiazarian, K.; Katkovnik, V.; Mehta, R.; Jain, A. K.; Agarwalla, A.; Praveen, C. V. S.; Zhou, R.; Wen, H.; Zhu, C.; Xia, Z.; Wang, Z.; and Guo, Q. 2017a. NTIRE 2017 Challenge on Single Image Super-Resolution: Methods and Results. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 1110–1121.

Timofte, R.; Agustsson, E.; Van Gool, L.; Yang, M.-H.; Zhang, L.; Lim, B.; Son, S.; Kim, H.; Nah, S.; Lee, K. M.; et al. 2017b. Ntire 2017 challenge on single image super-resolution: Methods and results. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 1110–1121.

Wang, L.; Dong, X.; Wang, Y.; Ying, X.; Lin, Z.; An, W.; and Guo, Y. 2021a. Learning Sparse Masks for Efficient Image Super-Resolution. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.

Wang, X.; Xie, L.; Dong, C.; and Shan, Y. 2021b. Real-ESRGAN: Training Real-World Blind Super-Resolution with Pure Synthetic Data. In *IEEE International Conference on Computer Vision Workshops (ICCVW)*.

Wang, X.; Xie, L.; Dong, C.; and Shan, Y. 2021c. Real-ESRGAN: Training Real-World Blind Super-Resolution with Pure Synthetic Data. In *International Conference on Computer Vision Workshops (ICCVW)*.

Wang, X.; Yu, K.; Wu, S.; Gu, J.; Liu, Y.; Dong, C.; Loy, C. C.; Qiao, Y.; and Tang, X. 2018. ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks. In *European Conference on Computer Vision (ECCV)*, volume abs/1809.00219.

Xia, B.; Hang, Y.; Tian, Y.; Yang, W.; Liao, Q.; and Zhou, J. 2022. Efficient Non-Local Contrastive Attention for Image Super-Resolution. In *Association for the Advancement of Artificial Intelligence (AAAI)*.

Yan, B.; Bare, B.; Ma, C.; Li, K.; and Tan, W. 2019. Deep Objective Quality Assessment Driven Single Image Super-Resolution. *IEEE Transactions on Multimedia*, 21(11): 2957–2971.

Yang, L.; Zhang, Z.; Song, Y.; Hong, S.; Xu, R.; Zhao, Y.; Zhang, W.; Cui, B.; and Yang, M.-H. 2023. Diffusion Models: A Comprehensive Survey of Methods and Applications.

Yusuke Matsui, Y. A. A. F. T. O. T. Y., Kota Ito; and Aizawa, K. 2017. Sketch-based manga retrieval using manga109 dataset. 76(20): 2181121838.

Z. Fang, X. L. J. W. L. L., W. Dong*; and Shi, G. 2022. Uncertainty learning in kernel estimation for multi-stage blind image super-resolution. In *European Conference on Computer Vision (ECCV)*.

Zeyde, R.; Elad, M.; and Protter, M. 2010. On single image scale-up using sparse-representations. In *International conference on curves and surfaces*, 711–730. Springer.

Zhang, L.; Zhang, L.; Mou, X.; and Zhang, D. 2011. FSIM: A Feature Similarity Index for Image Quality Assessment. *IEEE transactions on Image Processing*, 20(8): 7414–7426.

Zhang, R.; Isola, P.; Efros, A. A.; Shechtman, E.; and Wang, O. 2018a. The Unreasonable Effectiveness of Deep Features as a Perceptual Metric. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.

Zhang, W.; Liu, Y.; Dong, C.; and Qiao, Y. 2019. RankSRGAN: Generative Adversarial Networks With Ranker for Image Super-Resolution. In *IEEE International Conference on Computer Vision (ICCV)*, 3096–3105.

Zhang, Y.; Li, K.; Li, K.; Wang, L.; Zhong, B.; and Fu, Y. 2018b. Image super-resolution using very deep residual channel attention networks. In *European Conference on Computer Vision (ECCV)*.

Zhou Wang; Bovik, A. C.; Sheikh, H. R.; and Simoncelli, E. P. 2004. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4): 600–612.