# Downscaled Representation Matters: Improving Image Rescaling with Collaborative Downscaled Images

Bingna Xu*, Yong Guo*, Luoqian Jiang, Mianjie Yu, Jian Chen†
South China University of Technology
sexbn@mail.scut.edu.cn, guoyongcs@gmail.com,
{seluoqianjiang,202030482362}@mail.scut.edu.cn, ellachen@scut.edu.cn

## Abstract

*Deep networks have achieved great success in image rescaling (IR) task that seeks to learn the optimal downscaled representations, i.e., low-resolution (LR) images, to reconstruct the original high-resolution (HR) images. Compared with super-resolution methods that consider a fixed downscaling scheme, e.g., bicubic, IR often achieves significantly better reconstruction performance thanks to the learned downscaled representations. This highlights the importance of a good downscaled representation. Existing IR methods mainly learn the downscaled representation by jointly optimizing the downscaling and upscaling models. Unlike them, we seek to improve the downscaled representation through a different and more direct way – directly optimizing the downscaled image itself instead of the down-/upscaling models. Consequently, we propose a* ***Hierarchical Collaborative Downscaling (HCD)*** *method that performs gradient descent w.r.t. the reconstruction loss in both HR and LR domains to improve the downscaled representations, so as to boost IR performance. Extensive experiments show that our HCD significantly improves the reconstruction performance both quantitatively and qualitatively. Particularly, we improve over popular IR methods by $>0.57$ dB PSNR on Set5. Moreover, we also highlight the flexibility of our HCD since it can generalize well across diverse image rescaling models. The code is available at https://github.com/xubingna/HCD.*

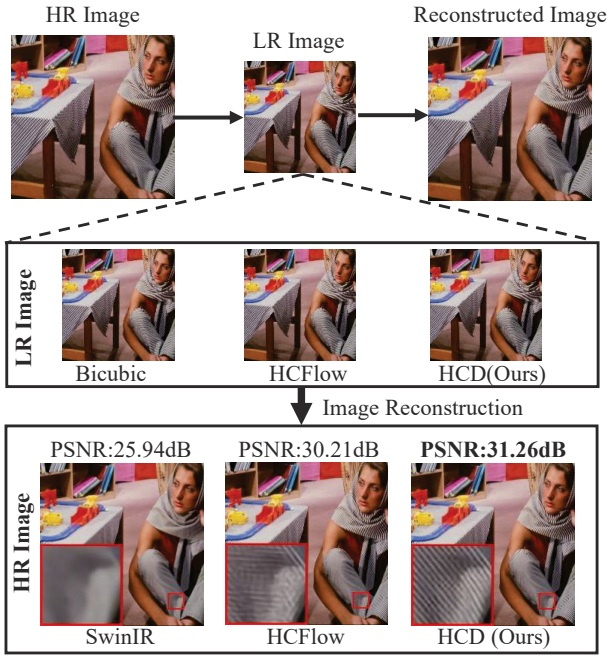


Figure 1. Image rescaling pipeline and the comparisons of the downscaled images along with the corresponding reconstructed HR images (4×). Top: we show the entire process of image rescaling. Middle: we visualize the downscaled representations used in different methods. Bottom: we compare the reconstructed HR images. With the improved downscaled representation, our method yields the best result both quantitatively and qualitatively.

## 1. Introduction

Image rescaling seeks to downscale the high-resolution (HR) images to visually valid low-resolution (LR) images and then upscale them to recover the original HR images. In practice, the downscaled images play an important role in saving storage or bandwidth and fitting the screens with different resolutions [43], such as image/video restoration and communication [48, 40, 35]. A typical application scenario of IR is to obtain HR images/videos (previously stored in the server) on an edge device, e.g., mobile. To save storage and reduce transmission latency, the original HR images/videos are usually downscaled to LR and then stored on the server. In some scenarios, these LR images can be directly used by edge devices, such as when the device screen has a low resolution or only as a preview, at the same time, they can also be upscaled to the original reso-

*Authors contributed equally.
†Corresponding author.

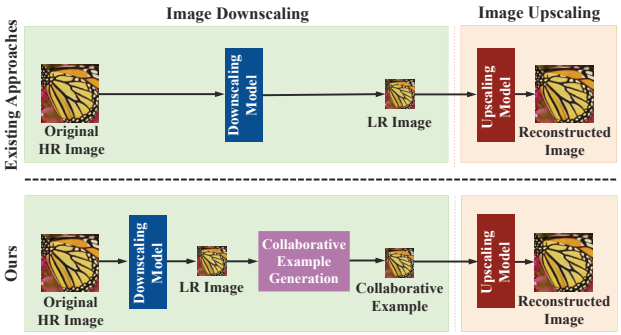


Figure 2. Comparison between existing image rescaling methods and the proposed HCD method. We additionally generate collaborative LR examples to improve the downscaled representation while keeping the upscaling process unchanged.

lution when needed. Interestingly, unlike super-resolution (SR) [38, 28, 3] methods that consider a fixed downscaling kernel, e.g., bicubic, IR often yields significantly better reconstruction performance [51, 4, 15] since IR essentially learns a better downscaled representation method. As shown in Figure 1, compared with a popular SR method SwinIR [25], the rescaling method HCFlow [26] greatly improves the PSNR score from 25.94 dB to 30.21 dB, which highlights the importance of a *good downscaled representation* in image reconstruction tasks.

To boost IR performance, existing methods jointly learn the downscaling and upscaling models by minimizing the reconstruction loss [43, 26]. However, in a complete pipeline, in addition to the trained neural network model, the downscaled representation itself is also very important. When diverse data are fed into a frozen model, we often obtain significantly different results. For example, in Figure 3, compared to the original LR images, the adversarial examples cause a 0.93 dB drop in PSNR of the reconstructed HR images, and visually, the lines became blurry. In contrast, when facing collaborative LR images, i.e., opposite to adversarial examples, not only improve the performance by 1.29 dB but also produce clearer and smoother lines. Inspired by this, in Figure 2, we propose a collaborative downscaling scheme that focuses on getting a *better downsampled representation* (purple box) of images instead of learning the model (blue box), making our approach essentially different from existing IR approaches. Furthermore, since LR images are obtained from the original HR image via downscaling, we can also improve the downscaled representation if we obtain a better representation in the HR domain, i.e., generating collaborative HR examples.

Motivated by this, we propose a **Hierarchical Collaborative Downscaling (HCD)** scheme that optimizes the representations in both HR and LR domains to obtain a better downscaled example. Specifically, we first generate a collaborative example in the HR domain and downscaled it to obtain an LR image. Taking the downscaled image as an initialization point, we then generate the collaborative LR example to further improve the downscaled representation. Due to the dependence between HR and LR image (based on the downscaling process), the hierarchical collaborative learning scheme can be formulated by a bi-level optimization problem. More critically, although our method increases the cost of generating downscaled images, we highlight that the increased cost ***only exists in the downscaling stage on server*** (can be processed offline), with no effect on the real-time rescaling on edge devices and making our method applicable to real-world scenarios. As shown in Figure 6, our HCD consistently improves PSNR while maintaining the same upscaling latency across diverse methods. Experiments show that our method significantly boosts the reconstruction performance with the help of collaborative downscaled examples.

Our contributions are summarized as follows:

- We propose a novel collaborative image downscaling method that improves the image rescaling performance from a new perspective – learning a better downscaled representation. We highlight that, in the community of image reconstruction, it is the first attempt to directly optimize the downscaled representation instead of learning the downscaling or upscaling models to boost the performance of image rescaling.

- Since the low-resolution (LR) images strongly depend on the corresponding high-resolution (HR) images, we propose a Hierarchical Collaborative Downscaling (HCD) that optimizes the representations in both HR and LR domains to learn a better downscaled representation. We formulate the learning process as a bi-level optimization problem and solve it by alternatively generating collaborative HR and LR examples.

- Experiments on multiple benchmark datasets show that, on top of state-of-the-art image rescaling models, our HCD yields significantly better results both quantitatively and qualitatively. For example, based on a strong baseline HCFlow [26], we obtain a large PSNR improvement of 0.7 dB on Set5 for 4× rescaling.

## 2. Related Work

### 2.1. Image Rescaling

Image rescaling (IR) and image super-resolution (SR) [41, 46, 47, 5, 10, 14] are distinct tasks. IR consists of image downscaling and image upscaling. SR corresponds to the latter process, which lacks a ground-truth HR image or any other prior information, and the reconstruction process is entirely based on the LR image. In contrast, we are given a ground-truth HR image in the IR task, but we use

its downscaled version for storage and transmission and recover the original HR image when necessary. Image downscaling is the inverse of SR which generates the LR version of an HR image. The most common downscaling method is bicubic interpolation [32], which may not be suitable for the upscaling tasks. However, it will cause over-smoothed issues since the high-frequency details are suppressed.

Recently, an increasing amount of work has been devoted to modeling image downscaling and upscaling as a unified task [9, 34, 23, 16, 50, 42]. Sun et al. [40] propose a content-adaptive resampler to achieve image downscaling and improves the upscaling model. Xiao et al. [43] propose a bidirectional image rescaling method based on invertible neural networks, which decomposes the image into low-frequency and high-frequency information through wavelet transformation, as the input of the model. Pan et al. [35] achieve bidirectional rescaling of arbitrary images by joint optimization. Recently, Zhong et al. [50] presents a generative prior reciprocated invertible rescaling network for the image rescaling task with an extreme upscaling factor($64\times$), which embeds the high-resolution information into the invertible low-resolution image to generative prior to the process of downscaling.

The above methods elaborately design the model architectures to reconstruct better images. Our method differs from theirs by directly optimizing the primeval input of the IR task under the supervision of the ground-truth HR image.

### 2.2. Image Super-Resolution

Image super-resolution (SR) is a widely-used image upscaling method that refers to recovering HR images from existing LR images. It is widely used in many applications, such as object detection [7], face recognition [33], medical imaging [18], and surveillance security [37]. Existing SR methods can be divided into three categories: interpolation-based, reconstruction-based, and learning-based methods. With the help of deep learning techniques, some SR methods achieve advanced effects on learning powerful prior information [6, 48, 49, 20, 24]. Ledig et al. [20] first propose SRGAN using Generative Adversarial Nets (GAN) [8, 11] to solve the over-smoothed problem of the SR task. Zhang et al. [48] combine the channel attention mechanism with SR to improve the representation ability of the model. Li et al. [24] propose SRFBN to refine low-level information using high-level ones through feedback connections and learn better representations. Sun et al. [39] propose HPUN to enhance the high-level representations with low-frequency features. Specifically, HPUN employs a lightweight pixel-unshuffled downsampler on the input LR images to obtain downsampling features with minimal information loss. Different from HPUN, Our proposed HCD is a ***test-time optimization*** method to improve inference performance without model modification or training. Indeed, HCD focuses on the direct optimization of downscaled representations yielded by the model. HCD generalizes well to diverse models and consistently improves results (Table 1).

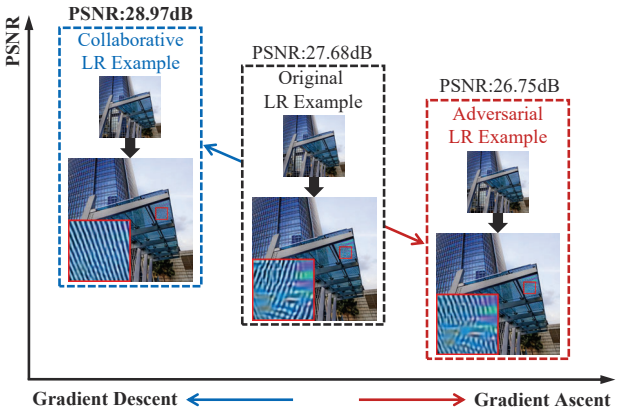


Figure 3. Comparison of adversarial examples and collaborative examples when producing downscaled representations. We ascend the gradient w.r.t the reconstruction loss to generate the adversarial examples. Based on the pretrained HCFlow model [26], the adversarial LR example yields lower PSNR along with distorted visual content in the reconstructed HR image. By contrast, we generate the collaborative LR example by descending the gradient and obtain significantly higher PSNR as well as better visual quality.

### 2.3. Collaborative Example

Generating collaborative examples [22] and adversarial examples [29, 12] are a set of opposite processes. The adversarial attack refers to the process of generating an adversarial example by applying a minor perturbation to the original input and causing the model to make an incorrect inference. The gradient-based attack method increases the prediction loss by updating the example along the positive direction of the gradient. In contrast to the adversarial example, the collaborative example aims to improve the robustness [22, 13] of the model by updating the example along the negative direction of the gradient, which ultimately decreases the prediction loss of the model. Inspired by the collaborative example, we propose a hierarchical collaborative example generation algorithm to generate downscaled representation optimal for the upscaling model under the supervision of the HR image.

## 3. Collaborative Image Downscaling

In this paper, we seek to directly learn a better downscaled representation rather than the down-/upscaling models to improve the performance of reconstructed images. In Section 3.1, we first discuss the importance of downscaled representation in image reconstruction tasks. Besides learn-

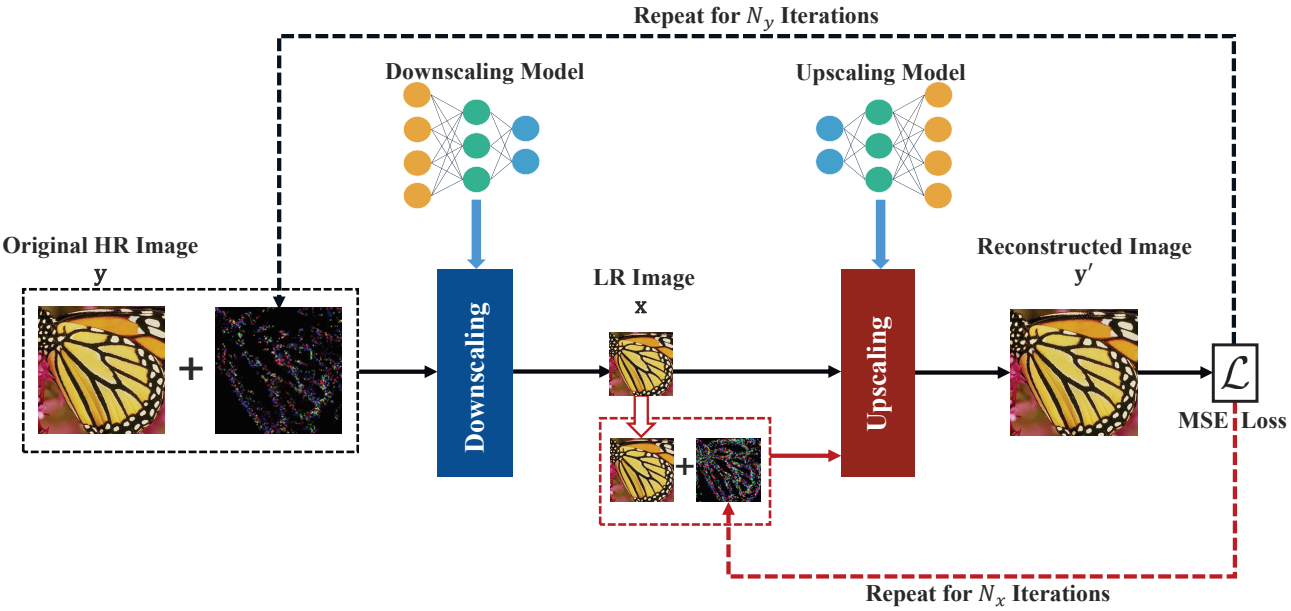


Figure 4. The proposed Hierarchical Collaborative Downscaling (HCD) scheme consists of two processes, including the collaborative HR example generation (marked as black lines) and the collaborative LR example generation (marked as red lines). We first iteratively optimize the perturbation on the original HR image to generate collaborative HR examples. Then, we obtain the downscaled image and generate collaborative LR examples for it. In the end, we can get a better high-resolution image from the better downscaled representation.

ing a good model, directly optimizing the downscaled representation (purple box in Figure 2) is an effective way to improve performance. In Section 3.2, we further extend this idea and propose a Hierarchical Collaborative Downscaling (HCD) method that optimizes the representations in both high-resolution (HR) and low-resolution (LR) domains to obtain a better downscaled representation. The overview of our HCD method is shown in Figure 4.

### 3.1. Downscaled Representation Matters

Existing image rescaling (IR) methods [43, 40, 35] essentially learn the optimal downscaled representation by jointly training the downscaling and upscaling models to reconstruct the original HR images. Compared with super-resolution methods that consider a fixed downscaling kernel, e.g., bicubic, IR methods often yield significantly better results thanks to the improved downscaled representation. For example, as shown in Figure 1, a recent rescaling method HCFlow [26] outperforms a strong baseline by a large margin of 5.27 dB. Such a large performance gap indicates the importance of a good downscaled representation in image reconstruction tasks.

In order to improve the downscaled representation, all the existing rescaling methods learn a downscaling method to produce the LR images. Besides training the model, one can also directly optimize the downscaled representation itself. From this perspective, a popular approach is adversarial attack [36] that learns the optimal perturbations on data without changing the model parameters. As shown by the red box in Figure 3, the adversarial attacks against image rescaling models greatly hamper the reconstruction performance in terms of both PSNR and visual quality. Nevertheless, we seek to improve the performance instead of degrading it. To address this issue, a simple and intuitive way is to generate collaborative examples by considering an opposite training objective to an adversarial attack. Specifically, we seek to generate collaborative examples in the LR domain by minimizing the reconstruction loss. As shown by the blue box, we can simultaneously improve PSNR and obtain visually plausible details in the reconstructed HR image. We highlight that the PSNR improvement of 0.93 dB is significant in image reconstruction tasks and generating collaborative examples provides us with new insights.

### 3.2. Hierarchical Collaborative Downscaling

In this part, we further extend the above idea and propose a novel Hierarchical Collaborative Downscaling (HCD) scheme to improve the downscaled representation, as shown in Figure 4. It is worth noting that, in theory, only optimizing LR images is possible to obtain the optimal reconstructed images (i.e., identical/close to HR images). But, in practice, we have to consider a limited number of gradient descent iterations to update LR images and constrain the perturbation $\delta$ with a $\epsilon$-ball. Given limited optimization budgets, the optimization results would heavily rely on the initial position of LR image $x$ on the manifold. Thus, it is

possible to obtain a better optimized LR based on a better initialization of $x$. To obtain better initialization, one can first optimize HR images. In this way, it becomes possible to obtain a better downscaled representation by generating collaborative examples in both the LR and HR domains. Essentially, jointly learning both collaborative examples can be regarded as a bi-level optimization problem. Algorithm 1 describes the pipeline.

To improve the reconstruction performance, we fix the model parameters and directly learn the optimal perturbation $\delta_y$ and $\delta_x$ to improve the downscaled representation $x$ sequentially. Let $\mathcal{L}$ be the reconstruction loss, $f(\cdot)$, and $g(\cdot)$ denote the upscaling and downscaling model, respectively. Following [29], we constrain the perturbation within a p-norm epsilon ball to avoid significantly changing the visual content, via $Clip\{\delta, \epsilon\} := \{\delta \mid \|\delta\|_p \leq \epsilon\}$. Without Clip, $\delta$ can be arbitrarily large and make the optimization very unstable. Formally, the perturbation in the LR domain for the bi-level optimization problem can be obtained by minimizing $\mathcal{L}$:

$$\begin{aligned} \delta_x &= \underset{\|\delta_x\|_p \leq \epsilon}{\arg\min}\, \mathcal{L}(f(g(y+\delta_y)+\delta_x), y), \\ \text{s.t. } \delta_y &= \underset{\|\delta_y\|_p \leq \epsilon}{\arg\min}\, \mathcal{L}(f(g(y+\delta_y)), y). \end{aligned} \tag{1}$$

As shown in Algorithm 1, we first generate collaborative HR examples and downscale them to obtain a better initialization of LR images $x$. Then, we further generate collaborative examples w.r.t. the latest LR images to obtain the resultant downscaled representation. Finally, the final LR images are fed into the upscaling model to produce the reconstructed images. In this way, we optimize the perturbations $\delta_x$ and $\delta_y$ in an iterative manner. For simplicity, we force both collaborative example generation processes to share the same number of iterations to perform gradient descent $N_x = N_y$. In this paper, we consider $\ell_2$-norm to build the epsilon ball, i.e., $p = 2$. Particularly, since we optimize the pipeline of IR tasks, the proposed HCD can be plug-and-play to be used on any advanced IR models to make performance better.

**Algorithm 1** Learning scheme of **Hierarchical Collaborative Downscaling (HCD)**. We present a hierarchical learning scheme that sequentially generates the collaborative HR examples and collaborative LR examples.

**Input:** HR image $y$, the downscaling model $g(\cdot)$, the upscaling model $f(\cdot)$, number of iterations $N_y$ and $N_x$, perturbation budget $\epsilon$, step size $\alpha$, the clipping function to constrain the input within feasible range $Clip\{\cdot\}$.
**Output:** Reconstructed high-resolution image $y'$.

Initialize the perturbations $\delta_y$ and $\delta_x$
*// Generate collaborative HR examples*
**for** $t = 1$ to $N_y$ **do**
    Compute gradient via $g = \nabla_{\delta_y} \mathcal{L}(f(g(y+\delta_y)), y)$
    Update $\delta_y$ via $\delta_y \leftarrow Clip\{\delta_y - \alpha * \frac{g}{\|g\|_p}, \epsilon\}$
**end for**
Obtain the collaborative HR example: $y = y + \delta_y$
Compute low resolution images: $x = g(y)$
*// Generate collaborative LR examples*
**for** $t = 1$ to $N_x$ **do**
    Obtain gradient via $g = \nabla_{\delta_x} \mathcal{L}(f(x+\delta_x), y)$
    Update $\delta_x$ via $\delta_x \leftarrow Clip\{\delta_x - \alpha * \frac{g}{\|g\|_p}, \epsilon\}$
**end for**
Obtain the collaborative LR example: $x = x + \delta_x$
Obtain the reconstructed image: $y' = f(x)$

# 4. Experiments

In the experiments, we evaluate the effectiveness of HCD based on three popular image rescaling methods, including IRN [43], HCFlow [26] and GRAIN [50]. We first describe the implementation details in Section 4.1. Then, we compare our method with current advanced methods in Section 4.2 on informative quantitative and qualitative analyses. Finally, we conduct abundant ablation studies and raise further discussions in Section 4.3. Both our source code and all the collaborative examples along with the corresponding reconstruction images will be released soon.

## 4.1. Implementation Details

We evaluate the proposed HCD on the validation set of DIV2K [1] and five standard datasets, i.e., Set5 [2], Set14 [45], BSD100 [30], Urban100 [17] and Manga109 [31]. We compare HCD with several state-of-the-art IR methods and SR methods. IR methods include TAD & TAU [19], CAR & EDSR [40], IRN [43], HCFlow [26] and an ultra-high rescaling method GRAIN [50]. However, SR with bicubic is also a possible solution to image rescaling and widely compared in IR papers [43, 26, 50, 15], so we also compared with the SR method SwinIR [25] and LTE-SwinIR [21].

Following [27], with regard to images represented in the YCbCr (Y, Cb, Cr) color space, we quantitatively evaluate the PSNR and SSIM [44] on the Y channel of them and test in 2× and 4× scale downscaling and reconstruction. The perturbation budget $\epsilon$ is set to 0.3 and the inner step size $\alpha$ is set to 20/255 for all experiments. By default, the iteration numbers for constructing collaborative LR images $N_x$ and HR images $N_y$ are set to the same i.e. $N_x = N_y = 15$.

## 4.2. Comparisons with State-of-the-arts

This section reports the performance of image reconstruction results on PSNR and SSIM. We consider two kinds of reconstruction methods as our baselines: IR methods and SR methods and experiment on the best-performing IR

| Method | Scale | Set5 | | Set14 | | BSD100 | | Urban100 | | DIV2K | | Manga109 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | PSNR | SSIM | PSNR | SSIM | PSNR | SSIM | PSNR | SSIM | PSNR | SSIM | PSNR | SSIM |
| Bicubic & Bicubic | 2× | 33.66 | 0.9299 | 30.24 | 0.8688 | 29.56 | 0.8431 | 26.88 | 0.8403 | 31.01 | 0.9393 | 30.82 | 0.9349 |
| Bicubic & SwinIR[25] | | 38.35 | 0.9620 | 34.14 | 0.9227 | 32.44 | 0.9030 | 33.40 | 0.9393 | – | – | 39.12 | 0.9783 |
| TAD & TAU[19] | | 38.46 | – | 35.52 | – | 36.68 | – | 35.03 | – | 39.01 | – | – | – |
| CAR & EDSR [40] | | 38.94 | 0.9611 | 33.86 | 0.9206 | 32.31 | 0.9012 | 32.76 | 0.9340 | 38.26 | 0.9599 | – | – |
| LTE[21] | | 38.32 | 0.9618 | 34.24 | 0.9235 | 32.43 | 0.9026 | 33.50 | 0.9406 | 36.74 | 0.9499 | – | – |
| LTE + HCD (ours) | | 39.14 | 0.9649 | 35.45 | 0.9342 | 33.18 | 0.9094 | 34.46 | 0.9412 | 37.09 | 0.9521 | – | – |
| IRN[43] | | 43.99 | 0.9870 | 40.79 | 0.9777 | 41.32 | 0.9876 | 39.92 | 0.9865 | 44.32 | 0.9908 | 43.68 | 0.9926 |
| IRN + HCD (ours) | | **44.67** | **0.9886** | **41.31** | **0.9791** | **41.82** | **0.9884** | **40.11** | **0.9868** | **44.75** | **0.9915** | **43.91** | **0.9928** |
| Bicubic & Bicubic | 4× | 28.42 | 0.8104 | 26.00 | 0.7027 | 25.96 | 0.6675 | 23.14 | 0.6577 | 26.66 | 0.8521 | 24.90 | 0.7876 |
| Bicubic & SwinIR[25] | | 32.72 | 0.9021 | 28.94 | 0.7914 | 27.83 | 0.7459 | 27.07 | 0.8164 | – | – | 30.92 | 0.9151 |
| CAR & EDSR[40] | | 33.88 | 0.9174 | 30.31 | 0.8382 | 29.15 | 0.8001 | 29.28 | 0.8711 | 32.82 | 0.8837 | – | – |
| TAD & TAU[19] | | 31.81 | – | 28.63 | – | 28.51 | – | 26.63 | – | 31.16 | – | – | – |
| LTE[21] | | 32.81 | 0.9025 | 29.05 | 0.7928 | 27.86 | 0.7466 | 27.24 | 0.8195 | 30.98 | 0.8498 | – | – |
| LTE + HCD (ours) | | 34.06 | 0.9155 | 30.33 | 0.7997 | 28.76 | 0.7689 | 27.98 | 0.8234 | 31.28 | 0.8525 | – | – |
| IRN[43] | | 36.19 | 0.9451 | 32.67 | 0.9015 | 31.64 | 0.8826 | 31.41 | 0.9157 | 35.07 | 0.9318 | 35.94 | 0.9616 |
| IRN + HCD (ours) | | 36.63 | 0.9488 | 33.21 | 0.9076 | 32.03 | 0.8894 | 31.76 | 0.9180 | 35.34 | 0.9351 | 36.44 | 0.9628 |
| HCFlow [26] | | 36.29 | 0.9468 | 33.02 | 0.9065 | 31.74 | 0.8864 | 31.62 | 0.9206 | 35.32 | 0.9346 | 36.51 | 0.9641 |
| HCFlow + HCD | | **36.99** | **0.9506** | **33.56** | **0.9116** | **32.22** | **0.8919** | **32.00** | **0.9231** | **35.48** | **0.9361** | **36.95** | **0.9651** |
| GRAIN [50] | 64× | 22.33 | 0.7718 | 19.96 | 0.6055 | 21.56 | 0.6128 | 16.77 | 0.3886 | 19.03 | 0.4901 | 16.30 | 0.4993 |
| GRAIN + HCD (ours) | | **23.81** | **0.7981** | **21.09** | **0.6321** | **22.81** | **0.6448** | **17.45** | **0.4099** | **19.88** | **0.5109** | **16.96** | **0.5190** |

Table 1. Quantitative evaluation results (PSNR / SSIM) of image reconstruction on benchmark datasets. LTE denotes LTE-SwinIR in the paper. The **black** values indicate the best result. The gray rows indicate the results of our backbone-based HCD, while the previous line of which indicates the results of the backbone methods. When the number of iterations N=15, the backbone methods with HCD improve PSNR and SSIM metrics on each benchmark dataset.

models with few iterations. It is worth mentioning that our HCD performs in the inference stage and the model parameters will not be changed during the iteration. We report the quantitative evaluation result of 4×, 2×, and 64× scale. Our HCD significantly outperforms the baseline.

| **Method** | **IRN** | **IRN+HCD** | **HCFlow** | **HCFlow+HCD** |
|---|---|---|---|---|
| FID↓ | 42.14 | **39.18 (-2.96)** | 36.10 | **33.42 (-2.68)** |
| LPIPS↓ | 0.1685 | **0.1578 (-0.0107)** | 0.1716 | **0.1613 (-0.0103)** |

Table 2. Quantitative evaluation results (FID / LPIPS) of image reconstruction on BSD100 for 4× scale. The **black** values indicate the best result. The results show our HCD can improve the visual quality with a lower FID and LPIPS.

**Quantitative results.** We summarize the quantitative comparison results of HCD and other advanced methods in Table 1. On all datasets, Our HCD significantly achieves better performance than previous state-of-the-art methods on PSNR and SSIM. Compared with the original model, HCD significantly improves the reconstruction of HR images with 15 iterations. For the 2× scale reconstructed images, Our HCD improves by 0.19-0.68 dB compared with IRN method. However, since the HCFLow model does not provide a pre-trained model for the 2× scale, we do not report the HCFlow-based HCD experimental results. For the 4× scale reconstructed images, our HCD improves by 0.16-0.7 dB compared with HCFlow and improves by 0.27-0.44 dB on IRN method. Similarly, even on ultra-high rescaling tasks, our HCD still can improve by 0.66-1.48 dB on the latest IR method GRAIN which more strongly demonstrates the robustness of our method. Besides, based on the state-of-the-art super-resolution method LTE-SwinIR, we also obtain significant improvements. In practice, when the proposed HCD is applied to large-resolution images e.g. the DIV2K dataset, we follow the "tile by tile" method in the test code of SwinIR [25], which divides images into patches to upscale. We can still get comparable results, PSNR improved by 0.43 dB at 2× scale in IRN. In addition to the previous comparisons, we also evaluated our proposed method HCD on the BSD100 for 4× scale in terms of the perceptual scores FID and LPIPS. The comparison with original approaches in Table 2 shows that our HCD improves the overall image quality as measured by FID and LPIPS. In general, our HCD significantly outperforms the baseline for PSNR and SSIM, and the perceptual scores such as LPIPS and FID, respectively, which shows excellent performance.

**Qualitative results.** We qualitatively evaluate our HCD by demonstrating details of the upscaled images. As shown in Figure 5, the results of HCD based on HCFlow achieve exhibit superior details and attractive visual quality. In the last set of Figure 5, our HCD alleviates unnatural colors in images from IRN and HCFlow. And it produces neater lines without bothersome horizontal lines compared with IRN. This demonstrates that our HCD significantly outperforms HCFlow and IRN visually. We leave more qualitative results in the appendix for spacing reasons.

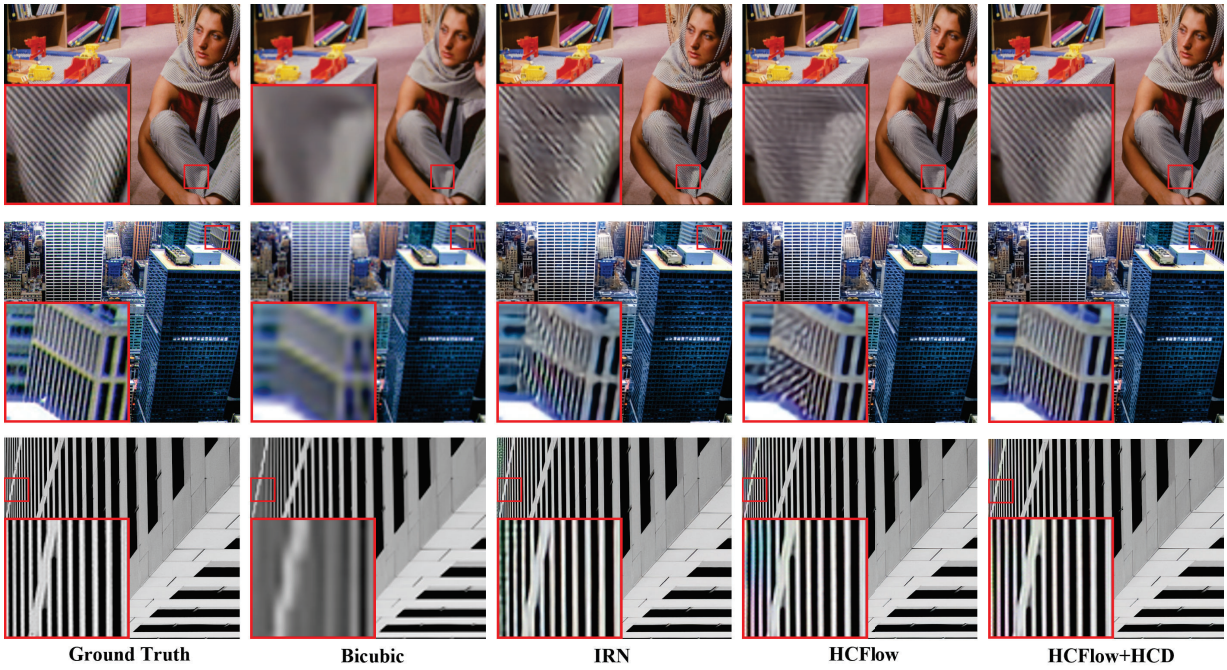


Figure 5. Qualitative results of upscaling the 4× downscaled images. Our HCD with HCFlow is able to produce more realistic and sharper HR images compared with the baseline methods. See the appendix for more results.

### 4.3. Ablations and Further Discussions

In this section, we present the results of the various iteration schemes and iteration numbers, as well as the effects of inner step size $\alpha$ and perturbation budget $\epsilon$. In addition to the reconstruction loss used in experiments, perceptual loss is also an important loss function in image tasks. Therefore, we also explore the influence of different loss functions on the proposed HCD. Besides, we discuss the impact of downscaling and upscaling latency. We also conduct ablations on HCD to justify each component. Unless otherwise specified, all ablation experiments are conducted on Set5 for $2\times$ scale based on IRN.

**Effect of the collaborative iteration number $N$.** In Table 3, we choose IRN as our backbone method and study the effect of different values of collaborative iteration number $N$ on building our collaborative example. In the default setting, we suggest $N_x = N_y = N$. We can find that the performance of IRN+HCD improves as the number of iterations increases, and only a few iterations are required to achieve a stable and satisfactory reconstruction effect.

**Effect of downscaling and upscaling latency.** Since the increased latency of our HCD only exists in the downscaling stage on the server, our HCD improves the performance of the backbone model without increasing the upscaling latency, as shown in Figure 6. There is no effect on the real-time upscaling on edge devices. In other words, we improve performance without increasing latency in reconstructing high-resolution images.

| # Iterations $N$ | PSNR | SSIM |
|---|---|---|
| 1 | 44.10 | 0.9872 |
| 5 | 44.52 | 0.9882 |
| 10 | 44.61 | 0.9884 |
| 15 | **44.67** | **0.9886** |
| 20 | 44.66 | 0.9872 |

Table 3. Effect of the number of iterations for generating collaborative examples ($N_x = N_y = N$) on the reconstruction performance. We report results on Set5 for $2\times$ rescaling based on IRN. Our HCD method gradually improves the reconstruction performance when we increase the collaborative iterations from 1 to 15. Nevertheless, the performance improvement becomes negligible if the iterations are increased to $N = 20$.

**Effect of hierarchical collaborative learning.** Table 4 reports the performance of different iteration combinations $(N_x, N_y)$ based on the backbone method IRN. When the iteration number degrades to zero, we skip the collaborative example generation step for the HR or LR images. When only collaborative LR examples or HR examples are used, the reconstructed image is improved by 0.47 dB and 0.42 dB after 15 iterations compared to IRN. When we leverage both examples, we can improve the performance by 0.21-0.26 dB over the best results achieved by these two examples alone.

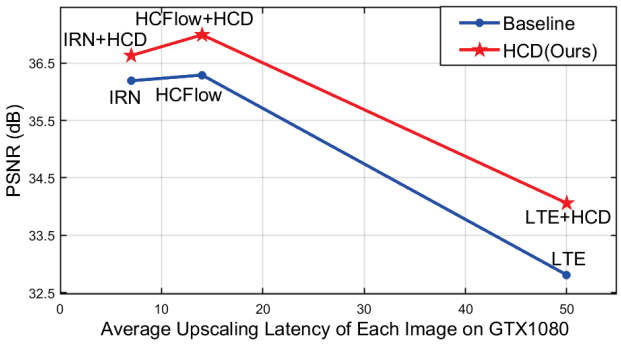


Figure 6. Comparison of upscaling latency and PSNR among different methods on Set5 for 4×. Just as the red line(our method) is always above the blue line(other IR methods), our HCD improves the performance of the backbone model without increasing the upscaling latency, which makes it applicable to real-world scenarios.

These results demonstrate the effectiveness of the proposed hierarchical learning scheme, showing that the collaborative HR examples can be combined with the collaborative LR examples to boost image reconstruction performance.

| # Iterations $N_x$ | # Iterations $N_y$ | PSNR | SSIM |
|---|---|---|---|
| 0 | 0 | 43.99 | 0.9870 |
| 30 | 0 | 44.46 | 0.9880 |
| 0 | 30 | 44.41 | 0.9880 |
| 15 | 15 | **44.67** | **0.9886** |

Table 4. The quantitative evaluation results (PSNR / SSIM) of different iteration schemes on Set5 at 2× scale based on IRN. Compared with optimizing LR or HR images separately(the second and third rows), the scheme of sequentially optimizing HR and LR images(the fourth row) increases PSNR by 0.21-0.26 dB.

**Effect of the step size $\alpha$.** In order to explore the effect of different inner step size $\alpha$, we keep the perturbation budget $\epsilon$ fixed at 0.3 and change the $\alpha$ in a range on IRN. The inference results are shown in Figure 7. Similar to the learning rate, a too-low $\alpha$ can slow down the inference process. However, since we randomly initialize the perturbations, a too-high $\alpha$ can also produce unreliable results. Experimental results show that our HCD performs well by varying $\alpha$ from 0.04 to 0.08.

**Effect of the perturbation budget $\epsilon$.** As shown in Figure 8, we indicate the effect of changing the perturbation budget on inference results. The maximum change in the image over the course of one iteration is related to the perturbation budget. As the figure illustrates, due to the limit of the change, a low perturbation budget can not produce a very good performance. The high perturbation budget allows for a wide range of pixel variations, resulting dramatically image changes, so it causes subpar performance.

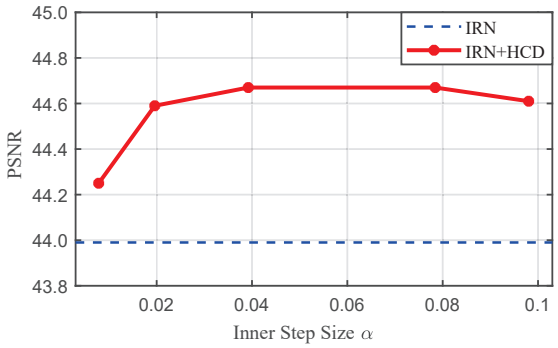


Figure 7. Experiment on choosing the different inner step sizes $\alpha$ on Set5 at 2× scale at the 15th iteration based on IRN. Our HCD performs well by varying the inner step size from 0.04 to 0.08.

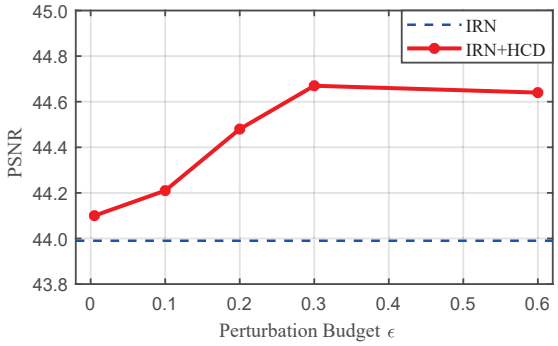


Figure 8. Experiment on choosing the different perturbation budget $\epsilon$ on Set5 at 2× scale at the 15th iteration based on IRN. The perturbation budget $\epsilon$ is expected to be as small as possible to bring less perturbation to make images visually imperceptible to change while maintaining performance. A moderate value of $\epsilon = 0.3$ brings a significant improvement.

We aim to achieve excellent performance improvement with minimal changes to the original images.

| Method | MSE Loss | Perceptual Loss | PSNR↑ | FID↓ |
|---|---|---|---|---|
| HCFlow | × | × | 36.29 | 32.15 |
| HCFlow+HCD | √ | × | **36.99** | 27.16 |
| | × | √ | 36.41 | **22.78** |
| | √ | √ | 36.75 | 24.57 |

Table 5. Quantitative results for different loss function on Set5 at 4× scale. Compared to MSE loss, the perceptual loss slightly reduces PSNR but greatly improves the visual quality(lower FID).

**Effect of the loss function $\mathcal{L}$.** In order to demonstrate the effect of the loss on HCD, experiments analyze the components in Eq. (1). As shown in Table 5, when only the MSE loss is used, the PSNR metric obtains the best results. While the perceptual loss will slightly reduce PSNR but greatly improves the visual quality of reconstructed images with a lower FID. These results indicate the flexibility of our HCD when combined with diverse losses.

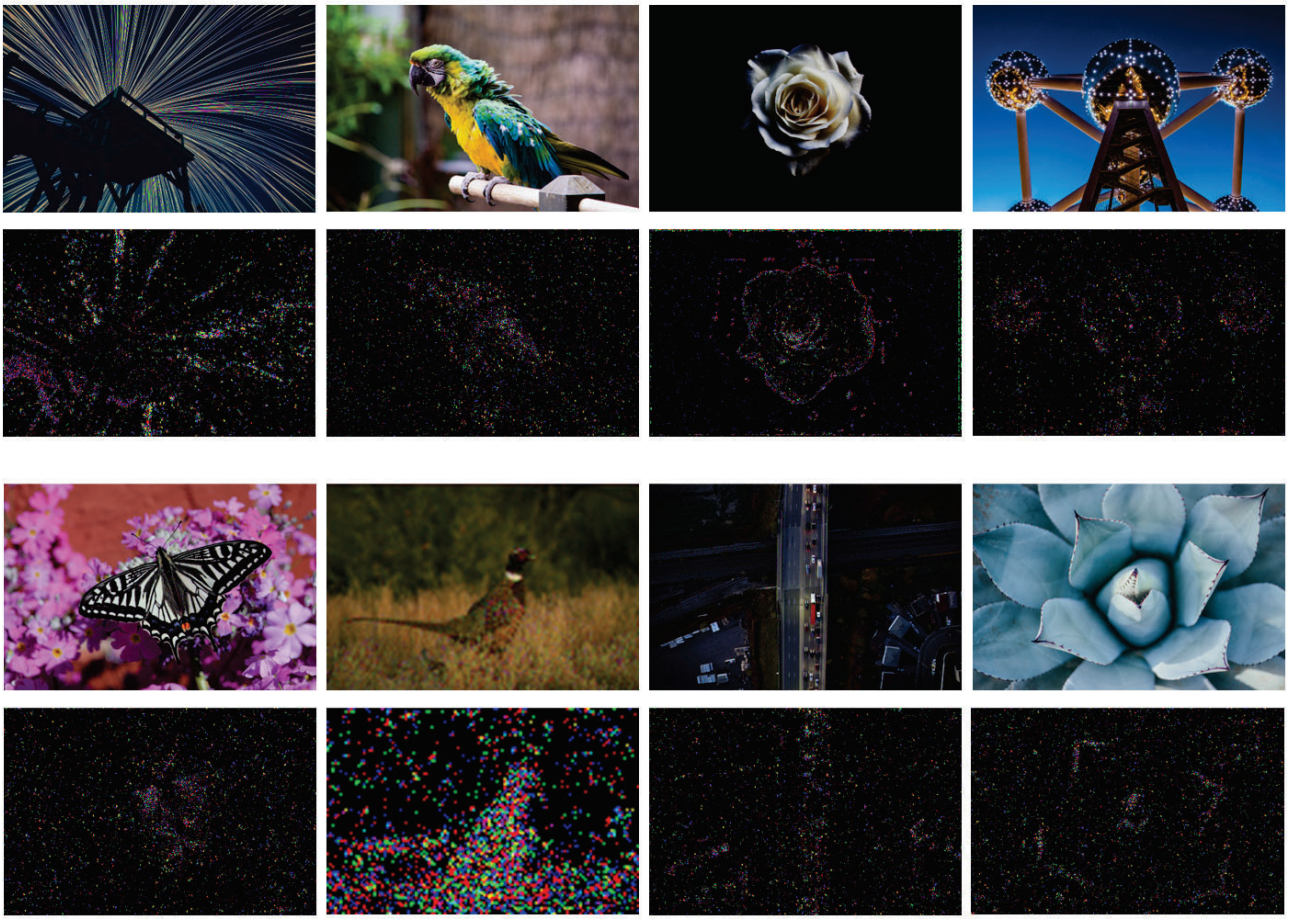

Figure 9. Visualization of the generated collaborative LR examples (top) and the corresponding perturbations $\delta_x$ (bottom). The perturbations added to the LR image are mainly distributed on the contours and corners of the image. These visualization results indicate that the generated collaborative examples are able to provide more information for those high-frequency regions, which, however, are often hard to be reconstructed in the upscaling process.

**Visualization of generated perturbations on LR images.** In Figure 9, we explicitly visualize the generated perturbations on the downscaled representation, i.e., $\delta_x$ in Eq. (1). The perturbations are mainly distributed on the contour and corner, such as the flower silhouette. Interestingly, these regions often contain high-frequency information that is hard to capture in the image upscaling process. We highlight that the performance improvement of our HCD method mainly stems from these collaborative perturbations.

## 5. Conclusion

In this paper, we propose a Hierarchical Collaborative Downscaling (HCD) method for the image rescaling task. In the first step, we generate collaborative samples for the input HR image of the downscaling model, so that it can be downscaled into a better LR representation. Then, we generate collaborative samples for the optimized LR to further improve its reconstruction performance. Extensive experiments show, both quantitatively and qualitatively, that our HCD significantly improves the performance on top of diverse image rescaling models.

## 6. Acknowledgments

This work was supported in part by the National Natural Science Foundation of China (Grant No. 62072186), in part by the Guangdong Basic and Applied Basic Research Foundation (Grant No. 2019B1515130001), and in part by the Opening Project of Guangdong Key Laboratory of Big Data Analysis and Processing.

## References


[1] Eirikur Agustsson and Radu Timofte. Ntire 2017 challenge on single image super-resolution: Dataset and study. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 126–135, 2017.

[2] Marco Bevilacqua, Aline Roumy, Christine Guillemot, and Marie-Line Alberi-Morel. Low-complexity single-image super-resolution based on nonnegative neighbor embedding. In *BMVC*, 2012.

[3] Honggang Chen, Xiaohai He, Linbo Qing, Yuanyuan Wu, Chao Ren, Ray E Sheriff, and Ce Zhu. Real-world single image super-resolution: A brief review. *Information Fusion*, 79:124–145, 2022.

[4] Yan-An Chen, Ching-Chun Hsiao, Wen-Hsiao Peng, and Ching-Chun Huang. Direct: Discrete image rescaling with enhancement from case-specific textures. In *2021 International Conference on Visual Communications and Image Processing (VCIP)*, pages 1–5. IEEE, 2021.

[5] Tao Dai, Jianrui Cai, Yongbing Zhang, Shu-Tao Xia, and Lei Zhang. Second-order attention network for single image super-resolution. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11065–11074, 2019.

[6] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Image super-resolution using deep convolutional networks. *IEEE transactions on pattern analysis and machine intelligence*, 38(2):295–307, 2015.

[7] Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. Region-based convolutional networks for accurate object detection and segmentation. *IEEE transactions on pattern analysis and machine intelligence*, 38(1):142–158, 2015.

[8] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks. *Communications of the ACM*, 63(11):139–144, 2020.

[9] Mengxi Guo, Shijie Zhao, Yue Li, Junlin Li, Li Zhang, and Yue Wang. Invertible single image rescaling via steganography. In *2022 IEEE International Conference on Multimedia and Expo (ICME)*, pages 1–6. IEEE, 2022.

[10] Yong Guo, Jian Chen, Jingdong Wang, Qi Chen, Jiezhang Cao, Zeshuai Deng, Yanwu Xu, and Mingkui Tan. Closed-loop matters: Dual regression networks for single image super-resolution. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5407–5416, 2020.

[11] Yong Guo, Qi Chen, Jian Chen, Qingyao Wu, Qinfeng Shi, and Mingkui Tan. Auto-embedding generative adversarial networks for high resolution image synthesis. *IEEE Transactions on Multimedia*, 21(11):2726–2737, 2019.

[12] Yong Guo, David Stutz, and Bernt Schiele. Improving robustness of vision transformers by reducing sensitivity to patch corruptions. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4108–4118, 2023.

[13] Yong Guo, David Stutz, and Bernt Schiele. Robustifying token attention for vision transformers. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023.

[14] Yong Guo, Jingdong Wang, Qi Chen, Jiezhang Cao, Zeshuai Deng, Yanwu Xu, Jian Chen, and Mingkui Tan. Towards lightweight super-resolution with dual regression learning. *arXiv preprint arXiv:2207.07929*, 2022.

[15] Kamali Gupta, Atul Garg, Vinay Kukreja, and Deepali Gupta. Rice diseases multi-classification: An image resizing deep learning approach. In *2021 International Conference on Decision Aid Sciences and Application (DASA)*, pages 170–175. IEEE, 2021.

[16] Xianxu Hou, Yuanhao Gong, Bozhi Liu, Ke Sun, Jingxin Liu, Bolei Xu, Jiang Duan, and Guoping Qiu. Learning based image transformation using convolutional neural networks. *IEEE Access*, 6:49779–49792, 2018.

[17] Jia-Bin Huang, Abhishek Singh, and Narendra Ahuja. Single image super-resolution from transformed self-exemplars. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5197–5206, 2015.

[18] Yawen Huang, Ling Shao, and Alejandro F Frangi. Simultaneous super-resolution and cross-modality synthesis of 3d medical images using weakly-supervised joint convolutional sparse coding. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 6070–6079, 2017.

[19] Heewon Kim, Myungsub Choi, Bee Lim, and Kyoung Mu Lee. Task-aware image downscaling. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 399–414, 2018.

[20] Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, et al. Photo-realistic single image super-resolution using a generative adversarial network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4681–4690, 2017.

[21] Jaewon Lee and Kyong Hwan Jin. Local texture estimator for implicit representation function. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1929–1938, 2022.

[22] Qizhang Li, Yiwen Guo, Wangmeng Zuo, and Hao Chen. Collaborative adversarial training. *arXiv preprint arXiv:2205.11156*, 2022.

[23] Shang Li, Guixuan Zhang, Zhengxiong Luo, Jie Liu, Zhi Zeng, and Shuwu Zhang. Approaching the limit of image rescaling via flow guidance. *arXiv preprint arXiv:2111.05133*, 2021.

[24] Zhen Li, Jinglei Yang, Zheng Liu, Xiaomin Yang, Gwanggil Jeon, and Wei Wu. Feedback network for image super-resolution. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 3867–3876, 2019.

[25] Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. Swinir: Image restoration using swin transformer. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1833–1844, 2021.

[26] Jingyun Liang, Andreas Lugmayr, K. Zhang, Martin Danelljan, Luc Van Gool, and Radu Timofte. Hierarchical conditional flow: A unified framework for image super-resolution and image rescaling. *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 4056–4065, 2021.

[27] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. Enhanced deep residual networks for single image super-resolution. *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1132–1140, 2017.

[28] Zhisheng Lu, Juncheng Li, Hong Liu, Chaoyan Huang, Linlin Zhang, and Tieyong Zeng. Transformer for single image super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 457–466, 2022.

[29] Aleksander Madry, Aleksandar Makelov, Ludwig Schmidt, Dimitris Tsipras, and Adrian Vladu. Towards deep learning models resistant to adversarial attacks. *arXiv preprint arXiv:1706.06083*, 2017.

[30] David R. Martin, Charless C. Fowlkes, Doron Tal, and Jitendra Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. *Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001*, 2:416–423 vol.2, 2001.

[31] Yusuke Matsui, Kota Ito, Yuji Aramaki, Azuma Fujimoto, Toru Ogawa, Toshihiko Yamasaki, and Kiyoharu Aizawa. Sketch-based manga retrieval using manga109 dataset. *Multimedia Tools and Applications*, 76:21811–21838, 2017.

[32] Don P Mitchell and Arun N Netravali. Reconstruction filters in computer-graphics. *ACM Siggraph Computer Graphics*, 22(4):221–228, 1988.

[33] Sivaram Prasad Mudunuri and Soma Biswas. Low resolution face recognition across variations in pose and illumination. *IEEE transactions on pattern analysis and machine intelligence*, 38(5):1034–1040, 2015.

[34] Zhihong Pan. Learning adjustable image rescaling with joint optimization of perception and distortion. In *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2455–2459. IEEE, 2022.

[35] Zhihong Pan, Baopu Li, Dongliang He, Mingde Yao, Wenhao Wu, Tianwei Lin, Xin Li, and Errui Ding. Towards bidirectional arbitrary image rescaling: Joint optimization and cycle idempotence. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 17389–17398, 2022.

[36] Shilin Qiu, Qihe Liu, Shijie Zhou, and Chunjiang Wu. Review of artificial intelligence adversarial attack and defense technologies. *Applied Sciences*, 9(5):909, 2019.

[37] Pejman Rasti, Tonis Uiboupin, Sergio Escalera, and Gholamreza Anbarjafari. Convolutional neural network super resolution for face recognition in surveillance monitoring. In *International conference on articulated motion and deformable objects*, pages 175–184. Springer, 2016.

[38] Chitwan Saharia, Jonathan Ho, William Chan, Tim Salimans, David J Fleet, and Mohammad Norouzi. Image super-resolution via iterative refinement. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(4):4713–4726, 2022.

[39] Bin Sun, Yulun Zhang, Songyao Jiang, and Yun Fu. Hybrid pixel-unshuffled network for lightweight image super-resolution. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 2375–2383, 2023.

[40] Wanjie Sun and Zhenzhong Chen. Learned image downscaling for upscaling using content adaptive resampler. *IEEE Transactions on Image Processing*, 29:4027–4040, 2020.

[41] Tong Tong, Gen Li, Xiejie Liu, and Qinquan Gao. Image super-resolution using dense skip connections. In *Proceedings of the IEEE international conference on computer vision*, pages 4799–4807, 2017.

[42] Qing Wang, Heidi R Howard, Juliana M Mcmillan, Guangxing Wang, and Xiaoyu Xu. A cnn-based rescaling algorithm and performance analysis for spatial resolution enhancement of landsat images. *International Journal of Remote Sensing*, 43(2):607–629, 2022.

[43] Mingqing Xiao, Shuxin Zheng, Chang Liu, Yaolong Wang, Di He, Guolin Ke, Jiang Bian, Zhouchen Lin, and Tie-Yan Liu. Invertible image rescaling. In *European Conference on Computer Vision*, pages 126–144. Springer, 2020.

[44] Hyunho Yeo, Youngmok Jung, Jaehong Kim, Jinwoo Shin, and Dongsu Han. Neural adaptive content-aware internet video delivery. In *OSDI*, 2018.

[45] Roman Zeyde, Michael Elad, and Matan Protter. On single image scale-up using sparse-representations. In *Curves and Surfaces*, 2010.

[46] Kai Zhang, Wangmeng Zuo, and Lei Zhang. Learning a single convolutional super-resolution network for multiple degradations. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3262–3271, 2018.

[47] Kai Zhang, Wangmeng Zuo, and Lei Zhang. Deep plug-and-play super-resolution for arbitrary blur kernels. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1671–1681, 2019.

[48] Yulun Zhang, Kunpeng Li, Kai Li, Lichen Wang, Bineng Zhong, and Yun Fu. Image super-resolution using very deep residual channel attention networks. In *Proceedings of the European conference on computer vision (ECCV)*, pages 286–301, 2018.

[49] Yulun Zhang, Yapeng Tian, Yu Kong, Bineng Zhong, and Yun Fu. Residual dense network for image super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2472–2481, 2018.

[50] Zhixuan Zhong, Liangyu Chai, Yang Zhou, Bailin Deng, Jia Pan, and Shengfeng He. Faithful extreme rescaling via generative prior reciprocated invertible representations. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5708–5717, 2022.

[51] Yiming Zhu, Cairong Wang, Chenyu Dong, Ke Zhang, Hongyang Gao, and Chun Yuan. High-frequency normalizing flow for image rescaling. *IEEE Transactions on Image Processing*, 2022.