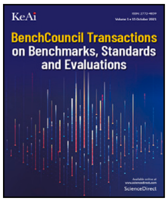
BenchCouncil Transactions on Benchmarks, Standards and Evaluations 3 (2023) 100148

Contents lists available at ScienceDirect

BenchCouncil Transactions on Benchmarks, Standards and Evaluations

journal homepage: www.keaipublishing.com/en/journals/benchcouncil-transactions-on-benchmarks-standards-and-evaluations/



Research article

A pluggable single-image super-resolution algorithm based on second-order gradient loss

Shuran Lin [a,b], Chunjie Zhang [a,b,*], Yanwu Yang [c]

[a] Beijing Key Laboratory of Advanced Information Science and Network Technology, Beijing Jiaotong University, Beijing, 100044, China
[b] Institute of Information Science, Beijing Jiaotong University, Beijing, 100044, China
[c] School of Management, Huazhong University of Science and Technology, Wuhan, 430074, China

ARTICLE INFO

Keywords:
Single-image super-resolution
Gradient
Texture
Loss function

ABSTRACT

Convolutional neural networks for single-image super-resolution have been widely used with great success. However, most of these methods use L1 loss to guide network optimization, resulting in blurry restored images with sharp edges smoothed. This is because L1 loss limits the optimization goal of the network to the statistical average of all solutions within the solution space of that task. To go beyond the L1 loss, this paper designs an image super-resolution algorithm based on second-order gradient loss. We impose additional constraints by considering the high-order gradient level of the image so that the network can focus on the recovery of fine details such as texture during the learning process. This helps to alleviate the problem of restored image texture over-smoothing to some extent. During network training, we extract the second-order gradient map of the generated image and the target image of the network by minimizing the distance between them, this guides the network to pay attention to the high-frequency detail information in the image and generate a high-resolution image with clearer edge and texture. Besides, the proposed loss function has good embeddability and can be easily integrated with existing image super-resolution networks. Experimental results show that the second-order gradient loss can significantly improve both Learned Perceptual Image Patch Similarity (LPIPS) and Frechet Inception Distance score (FID) performance over other image super-resolution deep learning models.

1. Introduction

As a well-known image restoration task, single-image super-resolution (SISR) aims to convert a low-resolution (LR) image into its corresponding high-resolution (HR) version. In recent years, SISR has gained significant attention from researchers owing to its practical applications in various fields, including video surveillance [1–3], medical imaging [4–6], and so on. Moreover, SISR can also be used in combination with other high-level computer vision tasks, such as object detection [7,8] and semantic segmentation [9,10], to improve their performance. However, SISR is inherently a challenging and ill-posed task, as LR images lack crucial texture details present in HR images, making it difficult to generate HR images from LR images alone. Furthermore, since a single LR image can be generated from multiple HR images that have undergone different types of degradation, the solution to the SR problem may not be unique.

Convolutional Neural Networks (CNNs) have recently shown impressive performance in information recovery due to their ability to handle complex data, and have thus been applied to the field of SISR. However, several CNN-based SISR methods currently popular prioritize high Peak Signal-to -Noise Ratio (PSNR) and Structural Similarity (SSIM) scores, which can lead to visually blurry restored images. This is because these methods often neglect the structural prior knowledge within the image and focus only on minimizing the mean absolute error between the recovered HR image and the ground truth image. Consequently, the optimization objective of the network becomes the statistical mean of all possible solutions in this one-to-many problem, resulting in blurry reconstructed images.

Images can be broken down into various frequency components, such as high-frequency and low-frequency components. The low-frequency component corresponds to its smooth regions, such as the sky, which are relatively simpler to restore. The high-frequency component pertains to its detailed regions, such as the textures of buildings, which are comparatively more challenging to restore. The human visual system is particularly sensitive to the details in images, especially the edges and textures, which play a crucial role in the perception of image quality [11]. Therefore, the accuracy of restoring the high-frequency

[*] Correspondence to: Beijing Key Laboratory of Advanced Information Science and Network Technology, Beijing Jiaotong University, No.3 Shangyuancun, Haidian District, Beijing, China.

E-mail address: cjzhang@bjtu.edu.cn (C. Zhang).

https://doi.org/10.1016/j.tbench.2023.100148
Received 16 August 2023; Received in revised form 8 October 2023; Accepted 8 December 2023
Available online 14 December 2023

components is essential for achieving visually pleasing results, and blurry images often occur when edge and texture details are lost due to excessive smoothing.

Numerous studies have demonstrated that incorporating prior knowledge of images, such as total variation prior [12,13], sparse prior [14–16], and gradient prior [17,18], can partially alleviate the ill-posedness of the SISR task. These prior knowledge can be viewed as supplementary constraints on the optimization objective of the network, which narrow down the solution space of the task. Among all these prior knowledge, the gradient prior is one of the most effective, as it can suppress noise and preserve edges during image reconstruction. In fact, an image can be regarded as a two-dimensional discrete function, and the gradient of the image is actually the derivative of this two-dimensional discrete function, which measures the change rate of the pixel grayscale value of the image. As the grayscale values of image pixels tend to vary greatly in edge and texture areas, the gradient map of images can accurately capture the edges and texture details of images. In the field of mathematics, the first-order derivative of a function provides information that can be utilized to describe the shape of the functional image, such as monotonicity. While the second-order derivative of the function contains more information than the first-order derivative, which has extremely important guiding significance for accurately modeling the functional image. Similarly, in the field of image processing, the second-order gradient map of images may contain more informative prior knowledge than the first-order gradient map. To validate this idea, we apply the principles of function derivation to generate the second-order gradient map of images and visualize it for a more intuitive comparison with the first-order gradient map. As shown in Fig. 1, the second-order gradient map shows more detailed information than the first-order gradient map. If fully utilized during network optimization, it can further compress the solution space of this task and reduce the difficulty of image restoration.

Based on the aforementioned discussion, this paper proposes an image super-resolution algorithm based on the second-order gradient (SG) loss. This algorithm replaces the loss function of the network with a combination of the SG loss and the L1 loss. The SG loss takes the second-order gradient map of the image as the starting point. To be specific, it first extracts the second-order gradient maps of the restored image and the HR image and then minimizes the distance between them to fully exploit the high-frequency information contained in the second-order gradient map of the image. This encourages the network to concentrate on the restoration of high-frequency components such as textures and image boundaries, improving the blurring of restored images caused by some existing methods that only use L1 loss as a constraint. The main contributions of this paper can be summarized as follows:

- We propose an image super-resolution algorithm based on the second-order gradient (SG) loss. By combining the SG loss with the L1 loss, our algorithm effectively guides the network optimization process and mitigates the problem of excessive blurring in the images restored by some existing image restoration methods to some extent. The SG loss can be easily integrated into most existing SR methods without adding extra training parameters.
- The experimental results on five widely used benchmark datasets demonstrate that the proposed SG loss can enhance high-frequency information in images and help the network recover clearer and more natural textures and edges.

2. Related works

This section provides a review of relevant image super-resolution methods from two perspectives: single-image super-resolution methods and gradient-guided super-resolution methods.

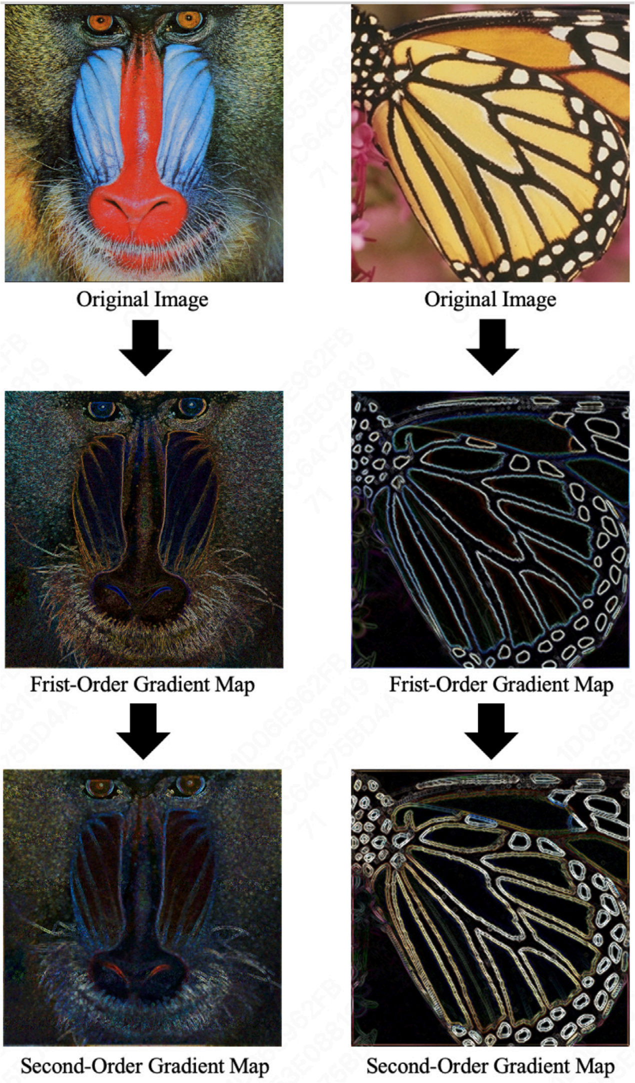


Fig. 1. Visualization of the first-order and second-order gradient maps of images.

2.1. Single image super-resolution methods

To date, numerous SISR methods have been proposed by researchers, which can be broadly classified into three categories: interpolation-based methods [19], signal processing-based methods [20–22], and deep learning-based methods [23–50].

In the initial stages of research on SISR, interpolation-based methods were commonly used. The main idea of these methods is to infer the pixel value at a specific position in the HR image by performing a weighted average of the known pixel values in the LR image surrounding that position. Different weighting schemes have been designed for image interpolation based on the fact that common pixel variations in a local region of an image can be approximated by a continuous function. For instance, bilinear interpolation, which leverages local linearity, and bicubic interpolation [19], which utilizes high-order continuity, are two examples of interpolation methods that have been proposed. Despite their simplicity and computational efficiency, these methods often result in generated images that exhibit unnatural artifacts and structural distortions. This is primarily due to the fact that the pixel variations within an image are often highly complex and cannot be accurately described by such simple predefined functions, particularly in the case of images with intricate textures. Signal processing-based SISR methods have been designed to address this issue. They apply signal processing techniques such as sparse representation [22], local adaptive filtering [21], and wavelet transform [20] to LR images to obtain their corresponding HR images. While signal processing-based methods have shown improvements in image restoration quality compared to interpolation-based methods, they often come with high computational complexity and are susceptible to noise.

For the last few years, deep learning-based methods have revealed extraordinary capabilities in feature learning and extraction, allowing neural networks to theoretically simulate any function. Through end-to-end model training, these deep learning networks can learn the mapping relationship between LR and HR images from massive data directly. These data-driven deep learning approaches lead to momentous performance gains compared to earlier traditional approaches. As trailblazers, Dong et al. are the first to establish a connection between CNN and image SR reconstruction. They devise a super-resolution convolutional neural network composed of three convolutional layers, which lay the groundwork for deep learning-based SISR methods. Nonetheless, the limited receptive field of the three convolutional layers restricts their capacity to perfectly leverage the surrounding pixel information, leading to constrained performance enhancement. For the purpose of enlarging the receptive field, Kim et al. [27] stack more convolutional layers and integrate residual learning to tackle the problem of gradient vanishing triggered by network thickening. Given the distinct sizes of LR and HR images in SISR, the aforementioned methods typically necessitate preprocessing of LR images utilizing bicubic interpolation to upscale them to match the size of HR images before feeding them into the network for training. Nevertheless, this preprocessing is time-consuming and exacerbates the noise and blur in LR images. To deal with this issue, a deconvolution layer is appended by Dong et al. [30] at the end of the network to accomplish end-to-end mapping from LR images to HR images. Shi et al. [32] present a novel sub-pixel convolutional layer that can achieve magnification by dynamically adjusting the number of feature channels. Both of them place the upscaling operation of the LR image at the final stage of the network and make it learnable. This can not only decrease the computational burden but also enhance the precision of image restoration.

Subsequently, SR models based on neural networks have emerged continuously. For instance, Zhang et al. [34] employ a dense connection structure to augment feature propagation through feature reuse. Li et al. [35] devise a multi-scale network to selectively extract image features of varying scales to facilitate image reconstruction, which leads to further performance improvement compared to the model using only a single scale. According to Zhang et al. [36], most existing methods treat LR input features indiscriminately and disregard the correlation between low-frequency information. Consequently, they integrate the attention mechanism into the SR network to enable it to concentrate on the more critical parts of the image for restoration. Several recent studies have attempted to combine transformers from the field of natural language processing with SR networks, obtaining state-of-the-art performance.

2.2. Gradient guided super-resolution methods

By exploiting gradient prior knowledge in many traditional methods [12,51–54], the solution space can be narrowed to generate a sharper image. For example, Fattal [52] designs a method that leverages image gradient edge statistics to learn the prior correlations across different resolutions. Zhu et al. [51] introduce an innovative method that gathers a dictionary of gradient patterns and characterizes deformable gradient combinations. Yan et al. [53] propose a stochastic resonance method based on gradient contour sharpness. Motivated by the effectiveness of gradient prior in traditional methods, some recent works have also endeavored to integrate image prior knowledge with neural networks [17,18,55]. Yang et al. [17] employ a pre-trained edge detector to extract image gradients, which are subsequently utilized to guide the deep network in reconstructing SR images. Ma et al. [18] construct a dual-branch joint optimization network consisting of a main SR branch and a gradient-assisted branch, where the gradient-assisted branch takes the gradient map extracted from the LR image as input, and the optimization target becomes the gradient map of its corresponding HR image. While previous methods leverage gradient prior knowledge to enhance the visual quality of restored images, they often incorporate learnable parameters associated with gradient information into the model significantly increasing its complexity and diminishing its computational efficiency. Unlike them, the proposed method in this paper utilizes a second-order gradient prior solely during network optimization to provide supplementary supervision information, without adding any learnable parameters. Hence, the computational cost can be disregarded.

3. Second-order gradient loss guided single-image super-resolution

3.1. Problem definition

For the task of SISR, the goal is to predict a reasonable HR image I^{SR} from a LR input image I^{LR}, given its corresponding ground truth HR image I^{HR}, and ensure that the predicted HR image I^{SR} is as similar as possible to the ground truth HR image I^{HR}. Consequently, during the actual model training, it is imperative to use pre-existing paired LR and HR image pairs (I^{LR}, I^{HR}). In reality, the LR image is typically obtained from the HR image through various types of degradation, but due to the complex and diverse forms of degradation and difficult modeling, for convenience of research, most works simply model the degradation process of the image as a bicubic interpolation downsampling operation. Therefore, the corresponding LR image can be generated from the HR image by the following formula:

$$I^{LR} = (I^{HR}) \downarrow_s \tag{1}$$

where $\downarrow_s$ denotes a bicubic interpolation downsampling operation with a scaling factor of s. Typically, both LR and HR images are 3-channel RGB images, with sizes of $3 \times h \times w$ and $3 \times s \cdot h \times s \cdot w$, respectively, where h and w are the height and width of the LR image. If we represent the SR network as F with parameters θ, then the process of image SR can be expressed as:

$$I^{SR} = F(I^{LR}; \theta) \tag{2}$$

Assuming that the loss function L can be applied to guide the network learning. In this case, we can formulate the optimization process of the network as follows:

$$\hat{\theta} = arg\min_{\theta} \mathbb{E}_{I^{SR}} L(F(I^{LR}; \theta), I^{HR}) \tag{3}$$

3.2. Second-order gradient loss

Most existing deep learning-based SR methods primarily rely on the L1 loss to constrain network training. The L1 loss is computed by measuring the mean absolute error between the predicted image I^{SR} generated by the network and the ground truth HR image I^{HR} at each pixel. This loss function tends to yield high Peak Signal-to-Noise Ratio (PSNR) values for the restored image. In fact, one of the limitations of using the L1 loss function in SR is that the visual results often exhibit blurriness and lack of preservation of sharp edges present in the original image. Despite the limitation aforementioned, the L1 loss function remains the most popular choice due to its effectiveness in accelerating convergence and improving the overall performance.

$$L_1 = \mathbb{E}_{I^{SR}} \|(I^{HR} - I^{SR})\|_1 \tag{4}$$

Considering that the L1 loss treats high-frequency and low-frequency information equally, without taking into account the fact that the inherent difficulty in recovering high-frequency details, this paper proposes to utilize the second-order gradient map of the image as additional

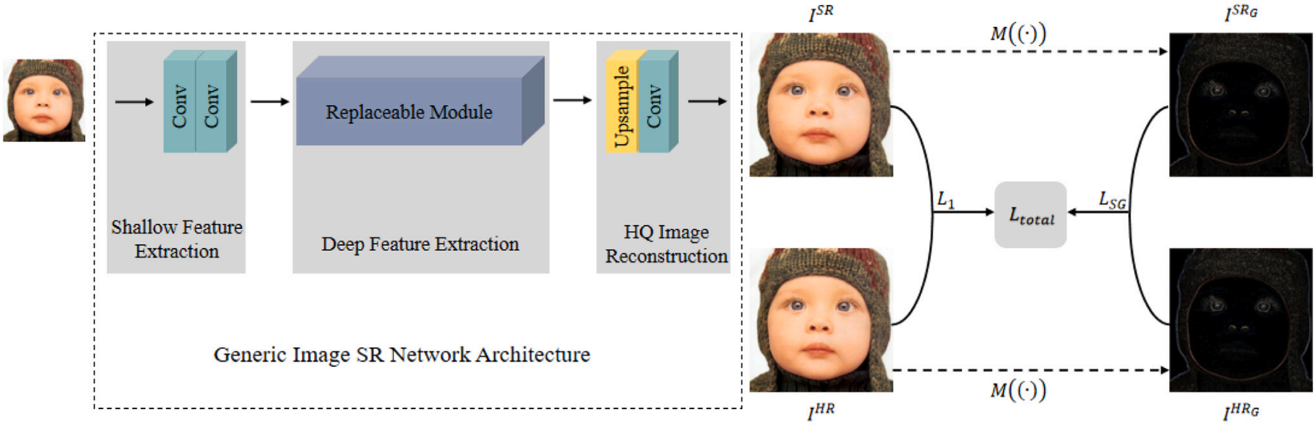


Fig. 2. Overall framework of our proposed pluggable SISR algorithm based on second-order gradient loss. The left half of this figure represents a generic deep learning-based image SR network architecture which can be easily replaced. I^{HR_G}, I^{SR_G} respectively represent the second-order gradient map extracted by using the gradient extraction function $M(\cdot)$ twice from high-resolution image I^{HR} and super-resolution image I^{SR}.

Table 1
Quantitative comparisons of cnn-based SISR models with and without second-order gradient loss on five benchmark datasets for ×4 SR. Best results are **highlighted**.

DataSet	Metric	EDSR	EDSR+SG	RDN	RDN+SG	RCAN	RCAN+SG	SwinIR	SwinIR+SG
Set5	LPIPS ↓	0.1728	**0.1446**	0.1716	**0.1560**	0.1720	**0.1401**	0.1700	**0.1412**
	FID ↓	58.86	**56.52**	57.88	**52.65**	59.74	**54.27**	58.80	**55.75**
Set14	LPIPS ↓	0.2776	**0.2353**	0.2808	**0.2564**	0.2783	**0.2268**	0.2705	**0.2262**
	FID ↓	86.45	**80.94**	88.75	**86.55**	91.95	**86.51**	89.17	**82.09**
Urban100	LPIPS ↓	0.2037	**0.1837**	0.2107	**0.1984**	0.2047	**0.1756**	0.1923	**0.1698**
	FID ↓	25.56	**23.10**	26.12	**23.85**	25.71	**22.39**	24.54	**21.63**
B100	LPIPS ↓	0.3589	**0.3018**	0.3634	**0.3274**	0.3602	**0.2906**	0.3549	**0.2894**
	FID ↓	96.08	**88.47**	96.36	**90.86**	98.15	**83.83**	95.59	**84.28**
Manga109	LPIPS ↓	0.0997	**0.0856**	0.1018	**0.0931**	0.0991	**0.0810**	0.0938	**0.0787**
	FID ↓	12.58	**10.77**	13.25	**11.39**	12.48	**10.80**	11.82	**9.97**

supervision information in the optimization process to encourage the network to pay more attention to high-frequency information during the recovery process and alleviate the problem of smoothing sharp edges. The reason why not utilizing higher-order gradient maps of the image is that studies have indicated that as the order of the gradient increases, the detail information in the gradient map becomes more intricate and complex, which may lead to instability during training and introduce additional errors. The loss function proposed in this paper involves the extraction of the second-order gradient map of the image. More precisely, to obtain the first-order gradient map of the image, we calculate the pixel-wise differences between adjacent pixels in both the horizontal and vertical directions. Subsequently, the second-order gradient map of the image is derived by calculating the pixel-wise differences between adjacent pixels of the first-order gradient map. During the actual training process, an additional constraint is imposed on the predicted high-resolution image I^{SR}. This constraint is to minimize the discrepancy between the second-order gradient maps of I^{SR} and I^{HR}. The gradient map of the image I can be generated using the following formula:

$$\begin{aligned} dx(i,j) &= I(i+1,j) - I(i-1,j) \\ dy(i,j) &= I(i,j+1) - I(i,j-1) \\ \nabla I(x,y) &= (dx(i,j), dy(i,j)) \\ M(I) &= \|\nabla I\| \end{aligned} \tag{5}$$

where (i,j) represents the coordinates of any point in the image, and $I(i,j)$ represents the pixel value of the image at (i,j). The operation $M(\cdot)$ refers to the process of extracting image gradients. It can be implemented by designing a convolution layer with fixed-weight kernels. In this paper, the weights of the convolution kernel are designed by simulating the Sobel filter. The Sobel filter is capable of detecting edge information in both the horizontal and vertical directions, making it an effective way to extract the gradient map from the image. Compared with other edge detection filters, it is not only simple to implement and fast in computation but also accurate in edge localization and good noise suppression in images. By applying the operation $M(\cdot)$ twice, we can obtain the second-order gradient map of the image. In summary, the proposed second-order gradient loss in this paper can be formulated as follows:

$$L_{SG} = \mathbb{E}_{I^{SR}} \| M(M(I^{HR})) - M(M(I^{SR})) \|_1 \tag{6}$$

Nevertheless, the second-order gradient loss primarily captures high-frequency information while lacking low-frequency information. To provide comprehensive guidance for network optimization, it is weighted and combined with the L1 loss to form the final loss function:

$$L_{total} = L_1 + \lambda L_{SG} \tag{7}$$

where λ is a hyperparameter that controls the weight of the SG loss L_{SG} in the total loss L_{total}.

To facilitate a more intuitive comprehension of the proposed second-order gradient loss, we have visualized it for illustrative purposes. As shown in Fig. 2, the left half of the figure represents a generic image SR network architecture, which consists of three parts: shallow feature extraction module, deep feature extraction module, and high-quality image reconstruction module. Typically, the shallow feature extraction module is composed of two convolutional layers that are intended to extract low-level features from the image and capture local information. The deep feature extraction module is more intricate and lacks a standardized structure, as it aims to extract high-level features from the image to capture global information. The high-quality image reconstruction module usually comprises an upsampling module and a convolutional layer. The upsampling module is responsible for increasing the size of the features, while the convolutional layer is responsible for transforming the features into an HR image. It is important to note that the specific architecture of the deep feature extraction module varies across different super-resolution networks, contributing to the variations in their restoration performance. The right half of the figure represents our proposed SG loss, it can be easily observed that it is a plug-and-play loss function, as the left half of the figure can be substituted with any image SR network. In summary, our proposed SG loss is generalizable. For ease of understanding, we list the meanings of the symbols used in this paper in Table 4.

4. Experiments

This section commences with an overview of the datasets, evaluation metrics, and implementation details employed in the experiment. Subsequently, a comparative analysis is presented, comparing the performance and visualization of several state-of-the-art SISR networks before and after the integration of the proposed SG loss. Furthermore, we investigate the influence of the λ value on the model performance.

4.1. DataSets and metrics

All the models are trained on the 800 images from the DIV2K [57] dataset. It is a widely recognized high-quality visual dataset in the field of SISR. For testing purposes, we utilize five standard public datasets: Set5 [58], Set14 [59], Urban100 [60], B100 [61], and Manga109 [62], which contain various scenes and can comprehensively analyze the effectiveness of the proposed loss. Since paired HR and LR images are required for training, the corresponding LR images are obtained by downsampling the HR images with a scaling factor of 4 using bicubic interpolation before conducting experiments. Considering that evaluation metrics such as Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity (SSIM) often contradict human perceptual quality, this paper adopts perceptual metrics Learned Perceptual Image Patch Similarity (LPIPS) [63] and Frechet Inception Distance score (FID) [64], which are more consistent with human perception, as evaluation metrics to quantitatively compare the restoration results of the datasets. Lower LPIPS and FID values indicate better visual quality.

4.2. Implementation details

For the purpose of conducting a fair comparison, several representative deep learning-based SR networks were retrained to establish a consistent benchmark. Specifically, during the training process, data augmentation techniques are performed on the training dataset. This includes random cropping, rotation by 90°, 180°, and 270°, as well as flipping the original images, resulting in approximately 32,000 HR images of size 480 × 480. In each training iteration, we take in 16 LR image patches with the size of 48 × 48 as input. The ADAM optimizer is employed for training, with default values of $\beta_1 = 0.9$, $\beta_2 = 0.999$ and $\epsilon = 1 \times 10^{-8}$. The learning rate is initialized as 1×10^{-4} and undergoes a halving operation every 2×10^5 iterations of back-propagation. Through empirical analysis, the hyperparameter λ is determined to be 1. Further details regarding the selection and impact of different λ values will be discussed in Section 4.5. The entire process is carried out on the PyTorch 2.0 platform, leveraging a Nvidia GeForce RTX 3090 24 GB GPU for accelerated computations.

4.3. Quantitative comparison

We select several widely recognized SR network models, including EDSR [56], RDN [34], RCAN [36], and SwinIR [37], to assess the effectiveness of our proposed SG loss function. No modifications have been made to their network architectures. Rather than using the L1 loss function alone, we augment it by adding an SG loss term to provide extra supervision information. The computational overhead incurred by this operation is negligible. The ×4 SR results on five benchmark datasets are presented in Table 1. The results marked with "+SG" indicate the outcomes obtained by adding the SG loss for auxiliary optimization to the original SR methods. The data in Table 1 demonstrates that the incorporation of the SG loss function as an auxiliary network optimization leads to lower LPIPS and FID values on all datasets for all models, compared to the original models. This observation strongly supports the belief that the second-order gradient map of the image, which contains high-frequency information, plays a crucial role in aiding the network to restore images with better perceptual quality. In particular, on the more severely degraded B100 dataset, our method achieves a substantial reduction in LPIPS scores for the SwinIR and RCAN models, with descents of 0.0655 and 0.0696, respectively. Additionally, the FID scores of these two models also exhibit notable improvements compared to the original method.

Table 2
Comparison of model restoration results trained with some training strategy and hyperparameters..

DataSet	Metric	1	2	3	4	5
Set5	LPIPS ↓	0.1446	0.1443	0.1447	0.1443	0.1441
	FID ↓	56.52	57.07	56.67	56.65	56.96
Set14	LPIPS ↓	0.2353	0.2361	0.2351	0.2355	0.2360
	FID ↓	80.94	81.54	80.89	80.05	81.36
Urban100	LPIPS ↓	0.1837	0.1837	0.1838	0.1836	0.1838
	FID ↓	23.10	22.92	23.08	22.95	22.87
B100	LPIPS ↓	0.3018	0.3018	0.3020	0.3023	0.3018
	FID ↓	88.47	88.64	88.03	87.63	87.32
Manga109	LPIPS ↓	0.0856	0.0856	0.0855	0.0856	0.0857
	FID ↓	10.77	10.67	10.65	10.70	10.80

4.4. Qualitative comparison

In order to provide further evidence of the effectiveness of our proposed SG loss, this section showcases visual results of the restored images obtained from the Set14, B100, and Urban100 datasets, with the majority of images selected from the Urban100 dataset. The Urban100 dataset was chosen for its collection of 100 images depicting buildings in urban areas. These images are rich in intricate texture details, making it an ideal dataset to demonstrate the effectiveness of the SG loss in restoring fine details. As illustrated in Fig. 3, the methods trained only using the L1 loss are capable of restoring the main contours of objects. However, they struggle to accurately restore complex image boundaries, often resulting in distorted and deformed textures. In contrast, after integrating the SG loss as supplementary supervision, the network preserves the fine details within images to a greater extent, and the reconstructed textures appear more natural and realistic.

4.5. Robustness experiment

To substantiate the robustness of the proposed SG loss function, we retrain the EDSR model multiple times using the same training strategy and hyperparameters, and the results of each training are exhibited in Table 2. We can find that although there are discrepancies in the recovery results of the models trained each time, these discrepancies are extremely minimal. This provides substantial evidence that the enhancement in model recovery performance attributed to the SG loss function is not incidental.

4.6. Ablation study of λ

To investigate the influence of different λ values on the performance of image restoration models, we train four distinct models with λ values of 0.01, 0.1, 1, and 10, respectively, employing the same training

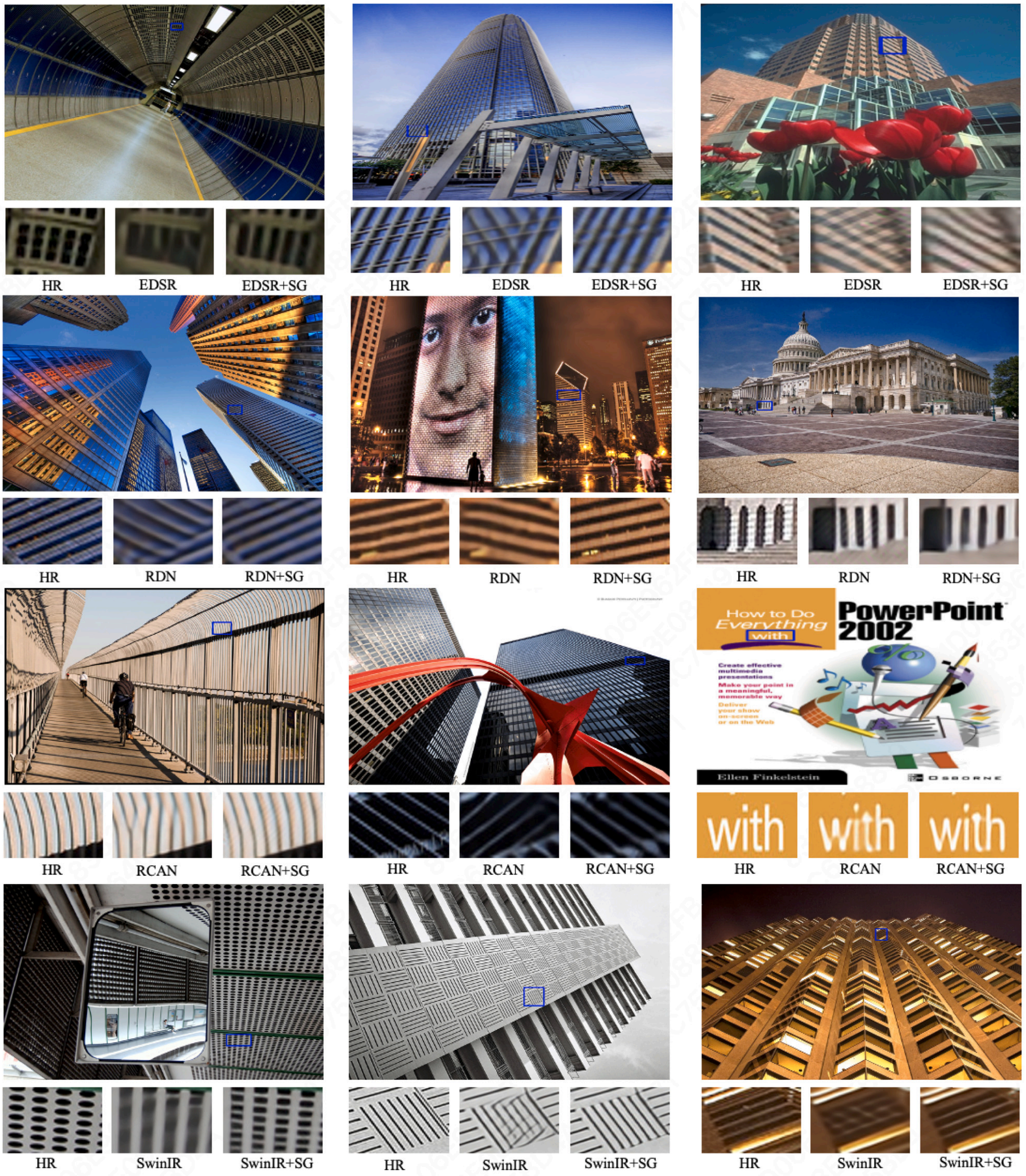


Fig. 3. Visual comparison of restoration results of different models before and after adding second gradient(SG) loss, where the first column of each image represents the GT high-resolution (HR) image, the second column represents the results recovered by EDSR [56], RDN [34], RCAN [36] and SwinIR [37], the third column represents the recovery results after integrating the SG loss by above methods.

Table 3
Comparison of model restoration results trained with different λ values.

Metrics	$\lambda = 0$	$\lambda = 0.01$	$\lambda = 0.1$	$\lambda = 1$	$\lambda = 10$
LPIPS ↓	0.2037	0.2013	0.1948	**0.1837**	0.1946
FID ↓	25.56	25.07	24.31	**23.10**	36.31

strategy. Subsequently, we conduct a comprehensive evaluation of the 4× SR performance of these four models on the Urban100 dataset. The results are presented in Table 3, with the highlighted numbers indicating the lowest LPIPS and FID scores in each row. To guarantee a fair comparison, all four models adopt the EDSR network structure. Fig. 4 is provided to offer a more intuitive observation of the differences in visual. The reference model, which does not incorporate the SG loss (i.e. $\lambda = 0$), exhibits the highest LPIPS and FID scores compared to the other models. As we increase the value of λ from 0 to 1, the

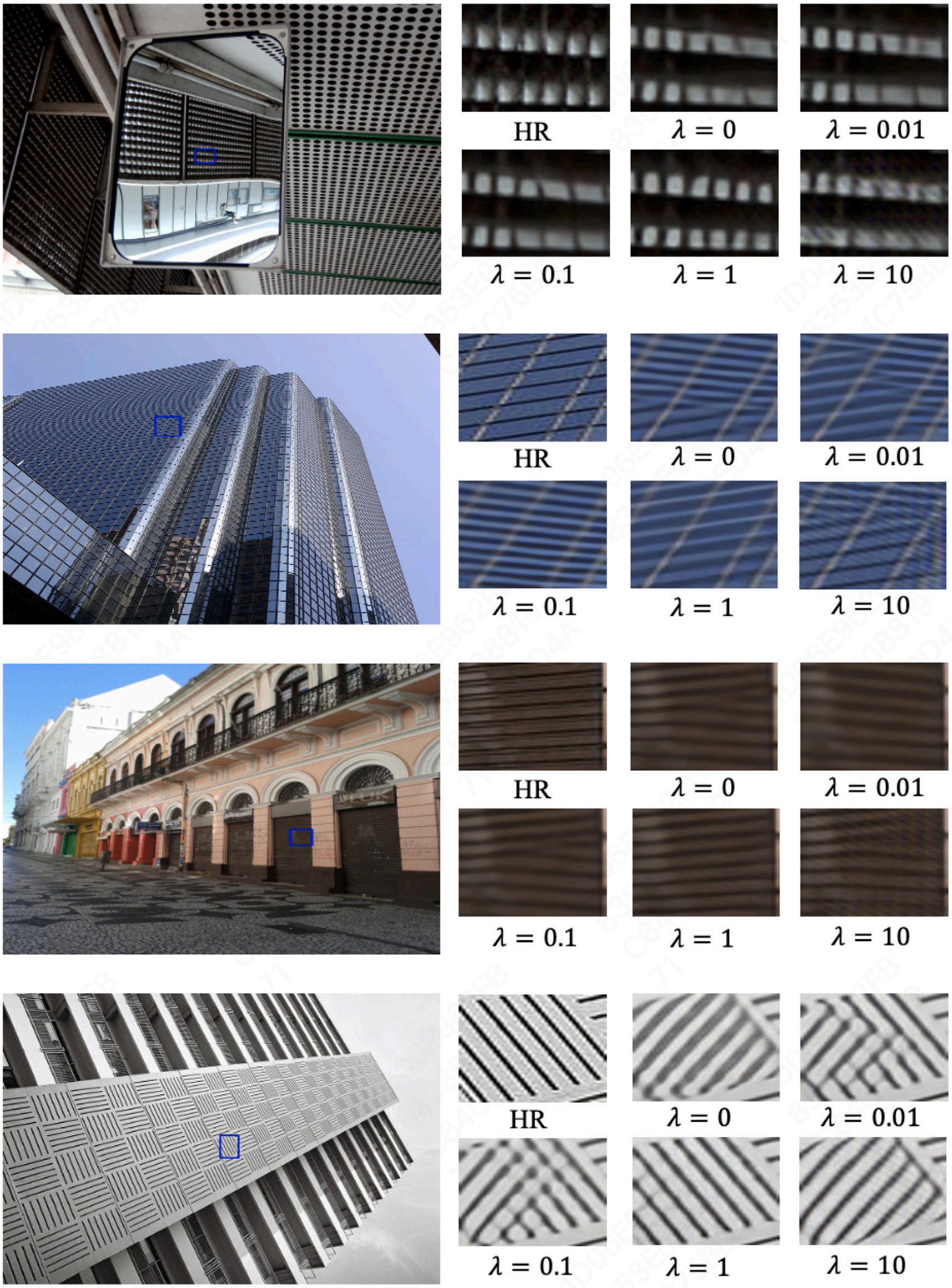


Fig. 4. Visual comparison of restoration results of models which trained with different λ values.

importance of the SG loss supervision gradually grows, resulting in a positive impact on the quality of the SR results. The best performance is attained when $\lambda = 1$. In this case, the model reaches a better balance between emphasizing image boundaries and preserving smooth regions. Nonetheless, if the λ value is excessively large, the model may overemphasize the textures and edges within images, and may even introduce unnatural artifacts in the smooth areas, resulting in a decline in model performance. Summarizing the above analysis, we recommend setting λ to 1 when using the SG loss.

4.7. *Ablation study of different loss functions*

We compare several loss functions commonly used in this field with our proposed SG loss function to further demonstrate its effectiveness, EDSR is still selected as the baseline for a fair comparison. Here, l2 loss [65] is a commonly used loss in the early days, charbonnier loss [66] is a variant of the l1 loss, which can better handle outliers and enhance model robustness, and ssim loss [67] can better simulate the perception of images by the human eyes. As shown in Table 5, when

Table 4
The meanings of symbols used in this paper..

Symbols	Meanings
I^{LR}	Low-Resolution Image.
I^{HR}	High-Resolution Image.
I^{SR}	Super-Resolution Image generated by our method.
$\downarrow_s$	Bicubic interpolation downsampling operation with a scaling factor of s.
I^{HR_G}	The second-order gradient map of the high-resolution image.
L^{SR_G}	The second-order gradient map of the super-resolution image generated by our method.
$dx(i,j)$	Horizontal gradient at the point (i, j).
$dy(i,j)$	Vertical gradient at the point (i, j).
$\nabla I(x,y)$	Horizontal and vertical gradient of image I.
$M(I)$	The first-order gradient map of image I.
$M(M(I))$	The second-order gradient map of image I.
L_1	L1 loss.
L_{SG}	Second-Order Gradient Loss.
L_{total}	The total loss used in this paper.
λ	The hyperparameter to balance the L1 loss and SG loss.

Table 5
Comparison of model restoration results trained with different loss functions.

Loss	LPIPS↓	FID↓
L1 loss	0.2037	25.56
L2 Loss	0.2064	25.04
SSIM loss	0.2057	24.98
Charbonnier Loss	0.2026	25.37
L1 loss + SG loss	**0.1837**	**23.10**

combined with the l1 loss function, our proposed SG loss can help the model achieve the best performance in all the quality metrics.

5. Conclusion

In this paper, an innovative high-frequency texture detail enhancement loss, referred to as the second-order gradient loss, is proposed to alleviate the problem of blurry high-resolution images generated by most existing SISR methods trained only with L1 loss. More specifically, the proposed second-order gradient loss function offers supplementary supervision for network optimization so that the solution space is compressed. This is accomplished by minimizing the discrepancy between the second-order gradient maps of the restored image and the high-resolution image. Furthermore, it can be seamlessly integrated with existing deep learning-based SISR methods without the need for introducing extra training parameters. This makes it a practical and convenient solution for enhancing the performance of SISR models without significant modifications to their existing architectures. The evaluation conducted on five public benchmark datasets indicate that the integration of this loss function significantly enhances the quality and fidelity of the restored images.

CRediT authorship contribution statement

Shuran Lin: Methodology, Software, Validation, Writing – original draft. **Chunjie Zhang:** Conceptualization, Formal analysis, Funding acquisition, Investigation, Methodology, Resources, Supervision, Validation, Writing – original draft, Writing – review & editing. **Yanwu Yang:** Data curation, Methodology, Resources, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work is supported by National Natural Science Foundation of China: 62072026, 72171093; Beijing Natural Science Foundation: JQ20022.

References

[1] L. Zhang, H. Zhang, H. Shen, P. Li, A super-resolution reconstruction algorithm for surveillance images, Signal Process. 90 (3) (2010) 848–859.

[2] Y. Pang, J. Cao, J. Wang, J. Han, JCS-net: Joint classification and super-resolution network for small-scale pedestrian detection in surveillance images, IEEE Trans. Inf. Forensics Secur. 14 (12) (2019) 3322–3331, http://dx.doi.org/10.1109/TIFS.2019.2916592.

[3] P. Shamsolmoali, M. Zareapoor, D.K. Jain, V.K. Jain, J. Yang, Deep convolution network for surveillance records super-resolution, Multimedia Tools Appl. 78 (2019) 23815–23829.

[4] Y. Huang, L. Shao, A.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, 2017, pp. 6070–6079.

[5] Y. Li, B. Sixou, F. Peyrin, A review of the deep learning methods for medical images super resolution problems, Irbm 42 (2) (2021) 120–133.

[6] D. Mahapatra, B. Bozorgtabar, R. Garnavi, Image super-resolution using progressive generative adversarial networks for medical image analysis, Comput. Med. Imaging Graph. 71 (2019) 30–39.

[7] Y. Bai, Y. Zhang, M. Ding, B. Ghanem, Sod-mtgan: Small object detection via multi-task generative adversarial network, in: Proceedings of the European Conference on Computer Vision, ECCV, 2018, pp. 206–221.

[8] J. Shermeyer, A. Van Etten, The effects of super-resolution on object detection performance in satellite imagery, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2019.

[9] L. Wang, D. Li, Y. Zhu, L. Tian, Y. Shan, Dual super-resolution learning for semantic segmentation, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 3774–3783.

[10] A. Aakerberg, A.S. Johansen, K. Nasrollahi, T.B. Moeslund, Semantic segmentation guided real-world super-resolution, in: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2022, pp. 449–458.

[11] S. Mandal, A.K. Sao, Edge preserving single image super resolution in sparse environment, in: 2013 IEEE International Conference on Image Processing, IEEE, 2013, pp. 967–971.

[12] M.K. Ng, H. Shen, S. Chaudhuri, A.C. Yau, Zoom-based super-resolution reconstruction approach using prior total variation, Opt. Eng. 46 (12) (2007) 127003.

[13] M.K. Ng, H. Shen, E.Y. Lam, L. Zhang, A total variation regularization based super-resolution reconstruction algorithm for digital video, EURASIP J. Adv. Signal Process. 2007 (2007) 1–16.

[14] J. Yang, J. Wright, T.S. Huang, Y. Ma, Image super-resolution via sparse representation, IEEE Trans. Image Process. 19 (11) (2010) 2861–2873.

[15] J. Yang, J. Wright, T. Huang, Y. Ma, Image super-resolution as sparse representation of raw image patches, in: 2008 IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2008, pp. 1–8.

[16] N. Akhtar, F. Shafait, A. Mian, Bayesian sparse representation for hyperspectral image super resolution, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 3631–3640.

[17] W. Yang, J. Feng, J. Yang, F. Zhao, J. Liu, Z. Guo, S. Yan, Deep edge guided recurrent residual learning for image super-resolution, IEEE Trans. Image Process. 26 (12) (2017) 5895–5907.

[18] C. Ma, Y. Rao, Y. Cheng, C. Chen, J. Lu, J. Zhou, Structure-preserving super resolution with gradient guidance, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 7769–7778.

[19] R. Keys, Cubic convolution interpolation for digital image processing, IEEE Trans. Acoust. Speech Signal Process. 29 (6) (1981) 1153–1160.

[20] D.K. Shin, Y.S. Moon, Super-resolution image reconstruction using wavelet based patch and discrete wavelet transform, J. Signal Process. Syst. 81 (2015) 71–81.

[21] A. Danielyan, A. Foi, V. Katkovnik, K. Egiazarian, P. Milanfar, Spatially adaptive filtering as regularization in inverse imaging: Compressive sensing super-resolution and upsampling, Super-Resolut. Imag. (2010) 123–154.

[22] K.I. Kim, Y. Kwon, Single-image super-resolution using sparse regression and natural image prior, IEEE Trans. Pattern Anal. Mach. Intell. 32 (6) (2010) 1127–1133.

[23] C. Dong, C.C. Loy, K. He, X. Tang, Image super-resolution using deep convolutional networks, IEEE Trans. Pattern Anal. Mach. Intell. 38 (2) (2015) 295–307.

[24] J. Li, F. Fang, J. Li, K. Mei, G. Zhang, MDCN: Multi-scale dense cross network for image super-resolution, IEEE Trans. Circuits Syst. Video Technol. 31 (7) (2021) 2547–2561, http://dx.doi.org/10.1109/TCSVT.2020.3027732.

[25] N. Ahn, B. Kang, K.-A. Sohn, Fast, accurate, and lightweight super-resolution with cascading residual network, in: Proceedings of the European Conference on Computer Vision, ECCV, 2018, pp. 252–268.

[26] M.S. Sajjadi, B. Scholkopf, M. Hirsch, Enhancenet: Single image super-resolution through automated texture synthesis, in: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 4491–4500.

[27] J. Kim, J.K. Lee, K.M. Lee, Accurate image super-resolution using very deep convolutional networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 1646–1654.

[28] Z. Zhang, Z. Wang, Z. Lin, H. Qi, Image super-resolution by neural texture transfer, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 7982–7991.

[29] J. Liu, W. Zhang, Y. Tang, J. Tang, G. Wu, Residual feature aggregation network for image super-resolution, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 2359–2368.

[30] C. Dong, C.C. Loy, X. Tang, Accelerating the super-resolution convolutional neural network, in: Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, the Netherlands, October 11-14, 2016, Proceedings, Part II 14, Springer, 2016, pp. 391–407.

[31] B. Niu, W. Wen, W. Ren, X. Zhang, L. Yang, S. Wang, K. Zhang, X. Cao, H. Shen, Single image super-resolution via a holistic attention network, in: Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XII 16, Springer, 2020, pp. 191–207.

[32] W. Shi, J. Caballero, F. Huszár, J. Totz, A.P. Aitken, R. Bishop, D. Rueckert, Z. Wang, Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 1874–1883.

[33] Y. Yan, W. Ren, X. Hu, K. Li, H. Shen, X. Cao, SRGAT: Single image super-resolution with graph attention network, IEEE Trans. Image Process. 30 (2021) 4905–4918, http://dx.doi.org/10.1109/TIP.2021.3077135.

[34] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, Y. Fu, Residual dense network for image super-resolution, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 2472–2481.

[35] J. Li, F. Fang, K. Mei, G. Zhang, Multi-scale residual network for image super-resolution, in: Proceedings of the European Conference on Computer Vision, ECCV, 2018, pp. 517–532.

[36] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, Y. Fu, Image super-resolution using very deep residual channel attention networks, in: Proceedings of the European Conference on Computer Vision, ECCV, 2018, pp. 286–301.

[37] J. Liang, J. Cao, G. Sun, K. Zhang, L. Van Gool, R. Timofte, Swinir: Image restoration using swin transformer, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 1833–1844.

[38] W.-S. Lai, J.-B. Huang, N. Ahuja, M.-H. Yang, Deep laplacian pyramid networks for fast and accurate super-resolution, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 624–632.

[39] K.-W. Hung, K. Wang, J. Jiang, Image interpolation using convolutional neural networks with deep recursive residual learning, Multimedia Tools Appl. 78 (2019) 22813–22831.

[40] T. Tong, G. Li, X. Liu, Q. Gao, Image super-resolution using dense skip connections, in: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 4799–4807.

[41] T. Dai, J. Cai, Y. Zhang, S.-T. Xia, L. Zhang, Second-order attention network for single image super-resolution, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 11065–11074.

[42] Y. Mei, Y. Fan, Y. Zhou, L. Huang, T.S. Huang, H. Shi, Image super-resolution with cross-scale non-local attention and exhaustive self-exemplars mining, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 5690–5699.

[43] Y. Zhang, D. Wei, C. Qin, H. Wang, H. Pfister, Y. Fu, Context reasoning attention network for image super-resolution, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 4278–4287.

[44] H. Chen, Y. Wang, T. Guo, C. Xu, Y. Deng, Z. Liu, S. Ma, C. Xu, C. Xu, W. Gao, Pre-trained image processing transformer, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 12299–12310.

[45] Y. Mei, Y. Fan, Y. Zhou, Image super-resolution with non-local sparse attention, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 3517–3526.

[46] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. 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, 2017, pp. 4681–4690.

[47] K. Zhang, W. Zuo, S. Gu, L. Zhang, Learning deep CNN denoiser prior for image restoration, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 3929–3938.

[48] J. Yu, Y. Fan, J. Yang, N. Xu, Z. Wang, X. Wang, T. Huang, Wide activation for efficient and accurate image super-resolution, 2018, arXiv preprint arXiv:1808.08718.

[49] M. Haris, G. Shakhnarovich, N. Ukita, Deep back-projection networks for super-resolution, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 1664–1673.

[50] Z. Hui, X. Wang, X. Gao, Fast and accurate single image super-resolution via information distillation network, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 723–731.

[51] Y. Zhu, Y. Zhang, B. Bonev, A.L. Yuille, Modeling deformable gradient compositions for single-image super-resolution, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 5417–5425.

[52] R. Fattal, Image upsampling via imposed edge statistics, in: ACM SIGGRAPH 2007 Papers, 2007, pp. 95–es.

[53] Q. Yan, Y. Xu, X. Yang, T.Q. Nguyen, Single image superresolution based on gradient profile sharpness, IEEE Trans. Image Process. 24 (10) (2015) 3187–3202.

[54] W. Dong, L. Zhang, G. Shi, X. Wu, Image deblurring and super-resolution by adaptive sparse domain selection and adaptive regularization, IEEE Trans. Image Process. 20 (7) (2011) 1838–1857.

[55] F. Fang, J. Li, T. Zeng, Soft-edge assisted network for single image super-resolution, IEEE Trans. Image Process. 29 (2020) 4656–4668.

[56] B. Lim, S. Son, H. Kim, S. Nah, K. Mu Lee, Enhanced deep residual networks for single image super-resolution, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2017, pp. 136–144.

[57] E. Agustsson, R. 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, 2017, pp. 126–135.

[58] M. Bevilacqua, A. Roumy, C. Guillemot, M.L. Alberi-Morel, Low-Complexity Single-Image Super-Resolution Based on Nonnegative Neighbor Embedding, BMVA Press, 2012.

[59] R. Zeyde, M. Elad, M. Protter, On single image scale-up using sparse-representations, in: Curves and Surfaces: 7th International Conference, Avignon, France, June 24-30, 2010, Revised Selected Papers 7, Springer, 2012, pp. 711–730.

[60] D.R. Martin, C.C. Fowlkes, D. Tal, J. Malik, A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics, in: Proceedings of the Eighth International Conference on Computer Vision, Vol. 2, ICCV-01, Vancouver, British Columbia, Canada, July 7-14, 2001, IEEE Computer Society, 2001, pp. 416–425, http://dx.doi.org/10.1109/ICCV.2001.937655.

[61] J.-B. Huang, A. Singh, N. Ahuja, Single image super-resolution from transformed self-exemplars, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 5197–5206.

[62] Y. Matsui, K. Ito, Y. Aramaki, A. Fujimoto, T. Ogawa, T. Yamasaki, K. Aizawa, Sketch-based manga retrieval using manga109 dataset, Multimedia Tools Appl. 76 (2017) 21811–21838.

[63] R. Zhang, P. Isola, A.A. Efros, E. Shechtman, O. Wang, The unreasonable effectiveness of deep features as a perceptual metric, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 586–595.

[64] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, S. Hochreiter, Gans trained by a two time-scale update rule converge to a local nash equilibrium, in: Advances in Neural Information Processing Systems, vol. 30, 2017.

[65] Z. Zhang, Parameter estimation techniques: A tutorial with application to conic fitting, Image Vision Comput. 15 (1) (1997) 59–76.

[66] P. Charbonnier, L. Blanc-Feraud, G. Aubert, M. Barlaud, Two deterministic half-quadratic regularization algorithms for computed imaging, in: Proceedings of 1st International Conference on Image Processing, Vol. 2, IEEE, 1994, pp. 168–172.

[67] Z. Wang, A. Bovik, H. Sheikh, E. Simoncelli, Image quality assessment: From error visibility to structural similarity, IEEE Trans. Image Process. 13 (4) (2004) 600–612, http://dx.doi.org/10.1109/TIP.2003.819861.