

RGSR: A two-step lossy JPG image super-resolution based on noise reduction



Biao Li ^{a,b,c,d}, Yong Shi ^{b,c,d,e,h}, Bo Wang ^f, Zhiqian Qi ^{b,c,d,*}, Jiabin Liu ^g

^a School of Business Administration, Southwestern University of Finance and Economics, Chengdu 611130, China

^b School of Economics and Management, University of Chinese Academy of Sciences, Beijing 101408, China

^c Research Center on Fictitious Economy and Data Science, Chinese Academy of Sciences, Beijing 100190, China

^d Key Laboratory of Big Data Mining and Knowledge Management, Chinese Academy of Sciences, Beijing 100190, China

^e College of Information Science and Technology, University of Nebraska, Omaha, NE 68182, USA

^f School of Information Technology and Management, University of International Business and Economics, Beijing 100029, China

^g Samsung Research China Beijing(SRC-B), Beijing 100028, China

^h College of Electronic & Information Engineering, Southwest Minzu University, Chengdu 610041, China

ARTICLE INFO

Article history:

Received 2 February 2020

Revised 30 July 2020

Accepted 27 August 2020

Available online 4 September 2020

Communicated by Ma Jiayi

Keywords:

Super-resolution

JPEG compression

Cycle loss

Image denoising

ABSTRACT

Single Image Super-Resolution (SISR) is a fundamental and important low-level computer vision (CV) task, yet its performance on real-world applications is not always satisfactory. Different from the previous SISR research, we focus on a specific but realistic SR issue: How can we obtain satisfied SR results from compressed JPG (C-JPG) images, which is a ubiquitous image format to greatly release storage space while missing fine details. the JPG SR task is deeply analyzed to discover the connotation. Then, we propose an effective two-step model structure named RGSR, involving two specifically designed components, i.e., JPG recovering and SR generation, instead of the perspective of noise elimination in traditional SR approaches. Besides, we further integrate the cycle loss to build a hybrid objective across scales for better SR generation. Experimental results on both of the standard test data sets and real images show that our approach achieves outstanding results and succeed in applying to practical C-JPG SR tasks.

© 2020 Elsevier B.V. All rights reserved.

1. Introduction

With the marvelous achievements of deep learning (DL) in computer vision (CV), super-resolution (SR) task attracts lots of attention for its crucial role as the basis of other high-level CV tasks, such as semantic segmentation [1] and image recognition [2]. Basically speaking, deep-learning super-resolution (DL-SR) algorithms, like [3–7], strive for simulating the map between low-resolution (LR) and its high-resolution (HR) counterpart through a complex nonlinear mapping network. In other words, the trained model approximates the inverse map of down-scaled operation used to obtain LR images from their HR fathers. This raises a troublesome issue: The success of this SR system highly hinges on the quality of LR images. If there are some spots/stains involving in LR inputs, the SR model will treat them as inherent elements of inputs, and the corresponding SR outputs will deservedly enlarge these undesirable details. We learn the issue from the application of SR models on cloth painting. The artifacts of pattern accompanied by C-JPG

enlarging are more obvious on cloth than paper. This phenomenon raises the barrier between the ideal SR model and practical use.

Most SR models are trained on the generated LR-HR dataset, like DIV2K [8]. Different from the ideal experimental setting, real-world images, e.g., pictures on the Internet, are usually compressed to reduce their sizes. However, compared with the lossless LR inputs, the quality of compressed JPG(C-JPG) images drop greatly, and further operations yield more unpleasant artifacts. For example, the presence of obvious partition lines coming from C-JPG vastly deteriorates the overall visual feeling. How to alleviate the deterioration makes the SR operation on C-JPG images to be a huge challenge. To address this issue, in this paper, we propose a specialized SR model: RGSR, which generate relatively satisfying high-resolution images from the low-quality inputs (C-JPG images).

Indeed, there are many SR pipelines designed for the real-world degradation SR tasks, like [9–16]. Some models regard the noise as a kernel estimation problem, solved by removing addictive Gaussian noises, such as [10,12]. However, the underlying realistic image distributions are usually inconsistent with the hypothetical Gaussian distribution, especially for the C-JPG issue in this paper which decreases information from the original image without

* Corresponding author at: School of Economics and Management, University of Chinese Academy of Sciences, Beijing 101408, China.

E-mail address: qizhiqian@foxmail.com (Z. Qi).

adding specific noises. Some models learn the corresponding maps by extracting useful representations from irrelevant LR-HR images, like the unsupervised learning strategy [13], the ensemble learning method [17]. Referring to the image compression artifact reduction, which is defined as part of image denoising, many techniques have been proposed, such as DnCNN [18], RDN [6], and RNAN[19]. These methods focus on how to recover details from the artifacts without concerning the practical application, like SR issue.

As a basic sense, training a DL network requires tremendous data. For the SR model, a large number of image pairs: LR images and their HR counterparts (served as the supervised information) are required for the network to learn the representation. In general, most LR images are produced through performing traditional interpolation methods (mostly Bicubic) on their HR fathers. As a result, the SR training process recovers this down-scaled mapping in a reverse manner. The relationship between the LR-HR pair with a fixed down-scaled kernel has been successfully learned by various SR models, such as SRGAN [20], EDSR [4], and RDN [6]. However, the repair mechanism of the C-JPG images includes more than SR generating process. When previewing the full-size image on the Internet, other than the down-scaled mapping, a number of additional unpleasant details are displayed, especially in the edges of objects. These compressed artifacts make former SR methods fail to generate applicable images. As shown in Fig. 1, the first four SR generations are trained on the LR-HR pairs related to the traditional Bicubic interpolation. Although the RCAN has been proved to be outstanding SR algorithms on a normal dataset, the results of C-JPG inputs still manifest our viewpoint that SR models trained with ideal LR-HR inputs will lead to poor SR performance. To be specified, the DPSR [10] specially designed a principled formulation and framework to handle LR images with arbitrary blur kernels. It generates unpleasant details in SR image. In detail, the damaged grids are apparently enlarged by the approaches designed for traditional non-JPG datasets. More specialized analysis can be found in [21].

A specially designed DL model - RGSR is proposed to address this drawback of the C-JPG SR task in this paper. The contributions of this paper can be summarized in the following three aspects. First, we deliberately introduce compressed JPG format LR images to construct a more complicated dataset with three types of training inputs: C-JPG LR, LR, and HR images. Second, the whole training process of RGSR includes two separate functional components: R (recovering stage) and S(super-resolution stage), i.e., the missing detail recuperative part (JPG recovering stage) and SR map learning part (SR generation stage). In order to remove the artifacts accompanied with C-JPG, such as ring and checkerboard, the first half sub-model R is trained with C-JPG LR images as its inputs, and leveraging the corresponding LR ones serve as its supervised information. Accordingly, its output named LR(C-JPG) removes the unpleasant noise and is greatly free of the partition lines phenomenon. Inherited these improved LR images, the latter

sub-model S continues to learn the map between LR(C-JPG) and HR images. Therefore, an SR generation pipeline between C-JPG and HR images is achieved through the joint integration of these two sub-models. Finally, because it has been repeatedly proved that multi-scale architecture can leverage complementary details in the image to contribute to better SR generations, a cycle loss across scales is plugged in the second stage. Experimental results demonstrate its effectiveness.

In short, the contributions of this paper can be summarized into three aspects:

- We firstly propose a specialized SR issue to utilize widespread image format, i.e., to generate satisfying and practical SR images from lossy JPG inputs, which bridges the main gap between the practical SR applications and the existing SR models.
- To address the proposed issue, we provide a two-step SR model named RGSR which involves two functional sub-models: the recovering part and the SR generating part. In detail, the JPG recovering stage firstly produces restored intermediates from C-JPG inputs, and the SR generation stage leverages the information of intermediates and HR images for the SR outputs. The experimental results demonstrate our approach can successfully solve the proposed task.
- To further improve the performance, we introduce a novel integrated loss function: two weighted L1 losses combining with a specific cycle-loss across multi-scale feature maps. We confirm the advantages of this loss function in our experiments.

2. Related work

In recent years, the concept of Big Data arises and many kinds of research, like [22–25], are done to explore the high dimensional data. Among them, tasks related to images, such as [26], attract a lot of attention for their wide applications.

Single image super-resolution (SISR) is a widely concerned task. A bundle of researches is developed to resolve this problem, and great achievements have been attained. In the beginning, researchers obtain SR generations with an estimated pixel-value interpolation function. For instance, Bicubic [27] explores the SR issue by the hypothesis that pixel spreading in a limited patch follows a specified linear parametric function. However, these traditional methods are prone to generate fuzzy SR results. Moreover, depending on the external or the internal dataset, we divide the SR methods into two types: external and internal based models. External SR models learn the mapping upon a large number of image pairs with various learning algorithms, such as nearest-neighbor [28], kernel ridge regression [29], sparse learning [30,31], cluster learning [32,33], manifold learning [34], and neural networks [35]. Meanwhile, internal SR models leverage information from similar patches within the image or across scales, and the approaches



Fig. 1. The comparing of SR generations coming from four methods: bicubic, RCAN, RCAN with Matlab pre-denosed input, and our model RGSR. The obvious different details are marked in red rectangles to indicate the contrast of four approaches.

mainly focus on how to obtain self-similar patches as much as possible [36–42]. In this part, we will briefly introduce some SR models based on the situation of training data.

2.1. Ideal condition images super-resolution

During the past several years, DL related methods have achieved excellent performance in many high-level CV tasks, as well as the SR field. These DL-SR methods focus on building high-efficient nonlinear models with tremendous parameters to modify the map between LR images and their corresponding HR fathers. To our knowledge, the most cutting-edge SR algorithms are all DL-dependent models in recent years. The phenomenon attribute to the powerful representation ability in the deep architectures. In detail, the first end-to-end DL-SR network is the SRCNN [35]. Though only three convolutional layers is included in the model, it greatly surpasses traditional SR methods in the accuracy metric of the peak-signal-to-noise-ratio (PSNR). Ever since, a number of novel DL-SR models are raised to pursue increasingly better performance, such as FSRCNN [43], VDSR [3], SRGAN [20], EDSR [4], RDN [6], DBPN [5], ESRGAN [44], EEGAN [45], and RCAN [7].

In general, DL-SR models are developed along with two different directions: higher accuracy (PSNR and SSIM values) and better photo-realistic visual sense (feature similarity). To design a model with better learning ability plays the most important role in this process. Based on this consideration, various structures are designed to extract more essential features from the LR image. For example, VDSR [3] introduces the global residual learning and trains a multi-scale model through sharing parameters across different scales. Inspired by the stunning performance of ResNet [2], the SRGAN [20] replaces the most basic convolutional blocks with residual ones to obtain more efficient performance. Besides, to achieve a better visual feeling, the SRGAN combines MSE loss, GANs loss [46] loss, perceptual loss [47], and global loss for photo-realistic SR generations. The effectiveness of new additional loss (i.e., perceptual loss and GANs loss) is proved by final photo-realistic SR results.

On the other hand, based on the generator of the SRGAN, the EDSR [4] introduces some changes: removing all batch normalization [48] layers, and replacing MSE loss with the $L1$ loss. Moreover, the number of channels is increased from 64 to 256. With these changes, EDSR contains over 40 million parameters and achieves state-of-the-art results in the NTIRE2017 [49]. Inspired by the Densely Network [50], the RDN [6] surpasses EDSR in accuracy (the PSNR score), on the condition of using less parameters. Then, the RCAN [7] upgrades the dense block to residual in residual (RIR) block and combines with the channel attention mechanism. To our knowledge, RCAN is one of the leading methods in accuracy pursuing SISR methods. The RefESR [17] pursuing better performance by an ensemble learning SR algorithm with a Reference dataset. The results prove the effectiveness of the introduced prior knowledge of ensemble weights learned from a reference dataset.

Meanwhile, another study aims to build a balanced pipeline between the model size and accuracy, such as CARN [51], s-LWSR [52], and FALSR [53]. These SR methods struggle for reducing parameters and operations while keeping their decent performance. For example, the CARN leverages the cascading mechanism upon a residual block to obtain better feature representation from multi-level layers. At the same time, in order to decrease the model size, some residual blocks are replaced by group convolution parts, which is similar to depth-wise convolution in MobileNets [54]. Besides, inspired by the U-Net [55], the s-LWSR introduces a more flexible SR model upon the U-Net structure with an additional information pool. Thanks to neural architecture search (NAS) [56], the FALSR automatically produces

a desirably lightweight and accurate SR model without human-design.

Despite these achievements, the above work is mostly built on the ideal LR-HR training pairs, which greatly limits their applications under the real-world conditions. Accordingly, building a more realistic SR model becomes the start point of our work.

2.2. Non-ideal condition image super-resolution

All methods mentioned above aim to solve the SR problem on supervised LR-HR image pairs. However, the scale-kernel and degradation function is usually undefined in the real-world, which results in the SR image accompanied by amplified noise. There are already many unsupervised SR models to tackle this issue.

For instance, the ZSSR [57] proposes an unsupervised Zero-shot SR approach. Firstly, the method generates many derived HR and LR images from the input, then trains a simple CNN network to learn the map between pre-proposed LR images and their HR counterparts. To our knowledge, ZSSR greatly surpasses other supervised SR models under the non-ideal conditions. Besides, based on CycleGAN [58], the CincGAN [13] proposes a cycle-in-cycle structure to address the blind SR problem. Inputs with noise are firstly processed to obtain intermediate LR images without these additional noises. Then, these intermediate LR images are jointly restored and scaled up with the help of an extra pre-trained SR model. To acquire SR images with desirable textures, the UMGSR [59] brings the multi-gram losses into the SR field. Combined with $L1$ loss, perceptual loss, and GAN-like adversarial loss, it generates visually satisfying images. For the supervised method, [60] proposes a novel pipeline to learn to generate SR images from the realistic data of the raw image. The EEGAN [45] tries to obtain the SR generations with high-frequency details from satellite images in noise-contaminated imaging conditions. Depends on the functional difference, the model is separated into two main subnetworks: an ultradense subnetwork (UDSN) and an edge-enhancement subnetwork (EESN).

In this paper, Our model shares a similar route leveraging multiple sub-models to solve the issue by steps. The non-ideal problem is related to the damage during the JPEG compression process which is in the widespread presence. Our model inherits a key prior assumption that the non-ideal SR issue can be gradually alleviated step by step.

3. Methodology

In this part, the content is arranged as follows: we first explain the challenge formulation. Then, the details of JPEG operation are presented to state how the information miss. Followed this, the architecture and loss function of the RGSR are described in detail.

3.1. Challenge formulation

Normally, the SISR issue is formulated as: $y = \mathcal{F}(x) + z$, where y and x represent HR images and LR images, respectively. Normally, the \mathcal{F} , e.g., a neural network, is the relationship between HR and LR in the hypothesis set space, while the z denotes some additional effects on the input, such as noises and blurred details. As mentioned in part 1, most DL-SR models are trained on the ideal dataset, where LR images are directly downsampled from HR fathers by a fixed kernel (e.g., Bicubic). In other words, the conversion changes the spreading of pixels with the linear transformation. Furthermore, we assume that the z equals to zero in an ideal dataset. However, these irrelevant information z accounts for the gap between ideal SR model and applications in reality.

In this paper, we investigate one special reality situation, where the LR image suffers loss of information from the low-quality JPEG compression. We call it the C-JPG SR issue. To be specified, this issue is ubiquitous. Because most images in reality, especially on the Internet, are stored in the C-JPG fashion to reduce the storage space or protect copyright. Since LR inputs are firstly deteriorated to low-quality images, the C-JPG SR issue can be redefined as: $y = \mathcal{F}(x + w)$. Here, w refers to the missing information due to the JPEG compression.

3.2. JPG operation

Image compression operation is the essential preprocessing to the minimum storage of images on the Internet. For example, an 800×800 image occupies as large as 1.7 M. In order to store it in a limited space, the image is processed by the JPEG technique with an alterable compression ratio. The detrimental compression operation mainly includes three steps: firstly, separating all data into important and unimportant zones; secondly, filtering the unimportant part; lastly, saving the rest as the output. The JPEG technique basically follows this procedure. We illustrate five primary processes in Fig. 2.

In detail, the JPEG process can be divided into following five steps.

(1) **Separating Blocks.** The input will firstly be divided into 8×8 blocks which will be further processed in the following steps.

(2) **RGB to YCbCr.** In this step, the color space will be transferred from the RGB area to the YCbCr zone. The RGB space represents that the color space of an image is represented with the red, the green, and the blue sub-areas, which is suitable for showing images on the display. For YCbCr, the Y, Cb, and Cr respectively represent the luminance, chromatic aberration of the blue, and the red color space. Because the luminance channel contains a lot of discardable details, the JPEG compression can deal with the input on this channel based on the importance.

(3) **Discrete Cosine Transformation.** The core of JPEG compression is the discrete cosine transform (DCT for short). With the help of the DCT, the hidden patterns in the data can be discovered from the chaos. More operations can be employed according to these patterns. Since the information of the small block is continuous, all patches' values normally are not greatly distinct

to each other. Specifically, the direct component in the top left corner stores the vast majority of information, while all other components are close to 0. Accordingly, all pixels are clearly separated into direct component and alternating components, which makes sufficient foreshadowing effect for the following compression.

(4) **Data Quantization.** Because information of the DCT processed data is still invertible. To free up more space, some details in the patch are filtered by data quantization. There are too many null values in the results of data quantization. These values can be controlled by a coefficient matrix, which is related to the quality of the JPEG result.

(5) **Huffman coding.** The Huffman coding [61] is the basis of the vast majority of compression algorithms. It is the standard pattern of the “ISO/IEC10918-1, 1993(e), Annex C”. To achieve a higher compression ratio, it determines the length of elements by their operating frequency. Huffman coding releases more space, 10 bytes for 64 arrays in the 8×8 patch.

To exhibit the difference, the comparison of LR(C-JPG) and LR(normal) is shown in Fig. 3. Given the original LR image, we firstly re-save it to its low-quality JPG versions as C-JPGs. It can be seen in Fig. 3 that the size of the original LR image reaches 95.7 K Bytes, while its C-JPG version, e.g., 20 percent saved in quality, only occupies 5.6 K Bytes. Accordingly, the quality of compressed images is greatly decreased with the increase of compressing ratio. This phenomenon can be clearly inferred from the visual contrast in Fig. 3, where mussy details appear in the C-JPG images, such as the irregular shape of windows and the blurry object edge. Normally, to achieve better balance between storage cost and human feeling satisfaction, many websites provide images with partial-quality JPG ones. When performing SR operation on these images, the effect of these deuterogenic noises is enlarged, leading to aggressively unpleasant feelings in visual. In this paper, we compress the LR images with 20 percent quality saved as the inputs, to better exhibit the great learning ability of our RGSR.

3.3. Network architecture

There are some two-step models, and each stage is defined by the special purpose, like the EEGAN [45] involving an ultradense subnetwork (UDSN) and an edge-enhancement subnetwork (EESN). Here, our work focuses on how to generate SR images from C-JPG inputs. Based on the specialization of the low-quality JPG-SR

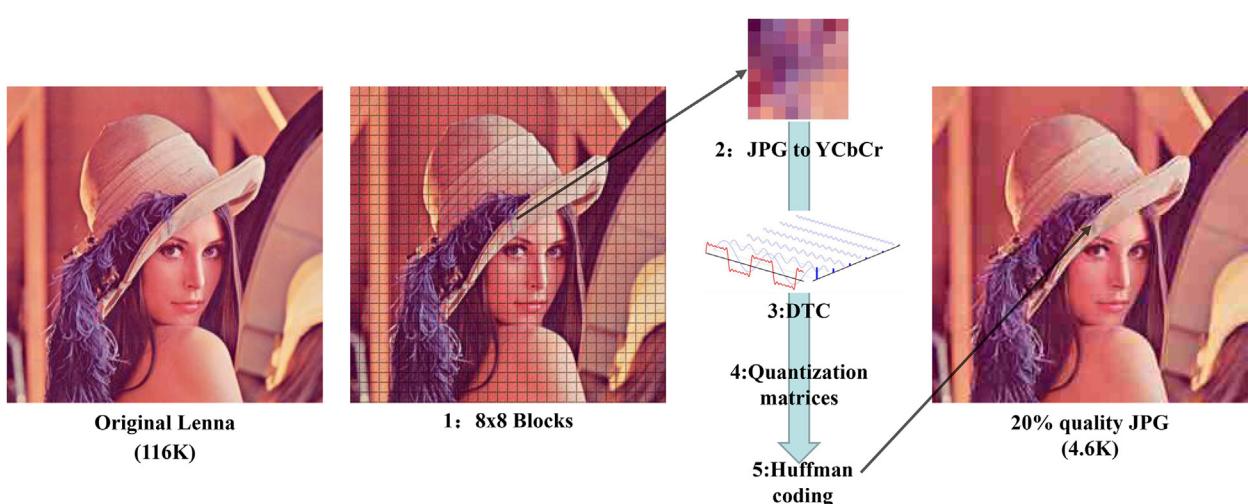


Fig. 2. The procedure of JPEG compression. We illustrate five primary steps. The quality of the input greatly drops down from 116 K to 4.6 K on condition that the size substantially shrinks.

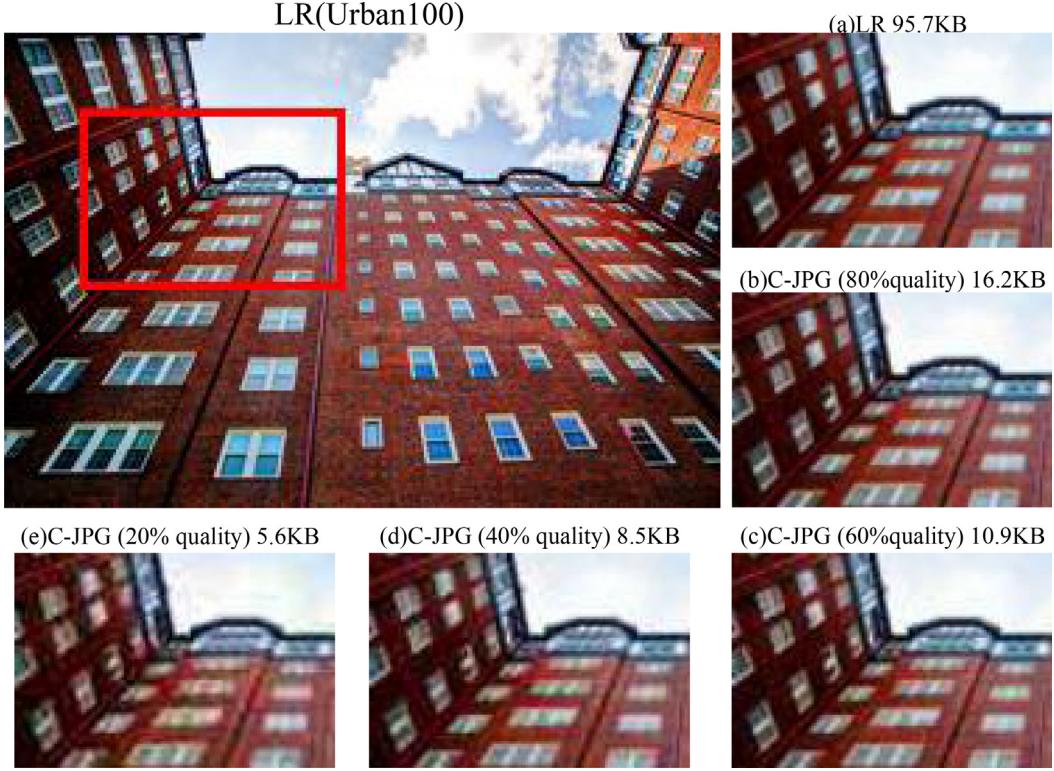


Fig. 3. The comparison of original LR image and its LR counterparts (obtained by different compressing ratios). We choose a typical building image from the Urban100 dataset [41]. Given the LR image, we leverage “Pillow” to compress it with four different quality indices: 80, 60, 40, and 20. Due to the compression operations, all C-JPG images occupy less storage space, from 95.7 KB to 5.6 KB. However, the visual feeling of C-JPG displays more unpleasant details with the increase of compressing ratio.

issue, our model can be separated into two functional stages: JPG recovering and SR generation. Hence, the RGSR formulation consists of two sub-stages:

$$I_{LR^\sim} = \mathcal{G}(I_{C-JPG}), \quad (1)$$

$$I_{SR} = \mathcal{F}(I_{LR}). \quad (2)$$

As shown in Fig. 4, three types of data are referred in the RGSR: C-JPG LR images ($LR(C-JPG)$); normal LR images ($LR(normal)$) downsampled from HR images, which provide supervised information in the first stage; and HR images, which are the supervised information in the second stage. The details are provided in the follows:

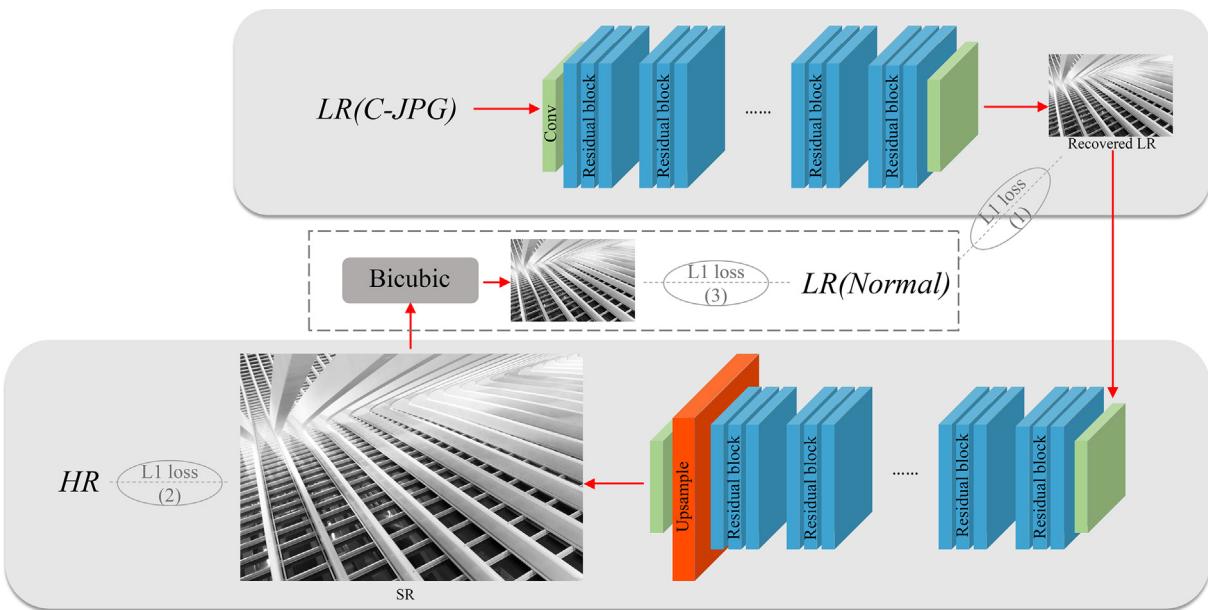


Fig. 4. The illustration of the proposed RGSR. The whole process can be divided into two sub-stages: the JPG recovering part and the SR generation part. Each stage is trained with a deep model. In this paper, we employ similar architectures for the very two tasks. The difference is the existence of the up-sampling stage.

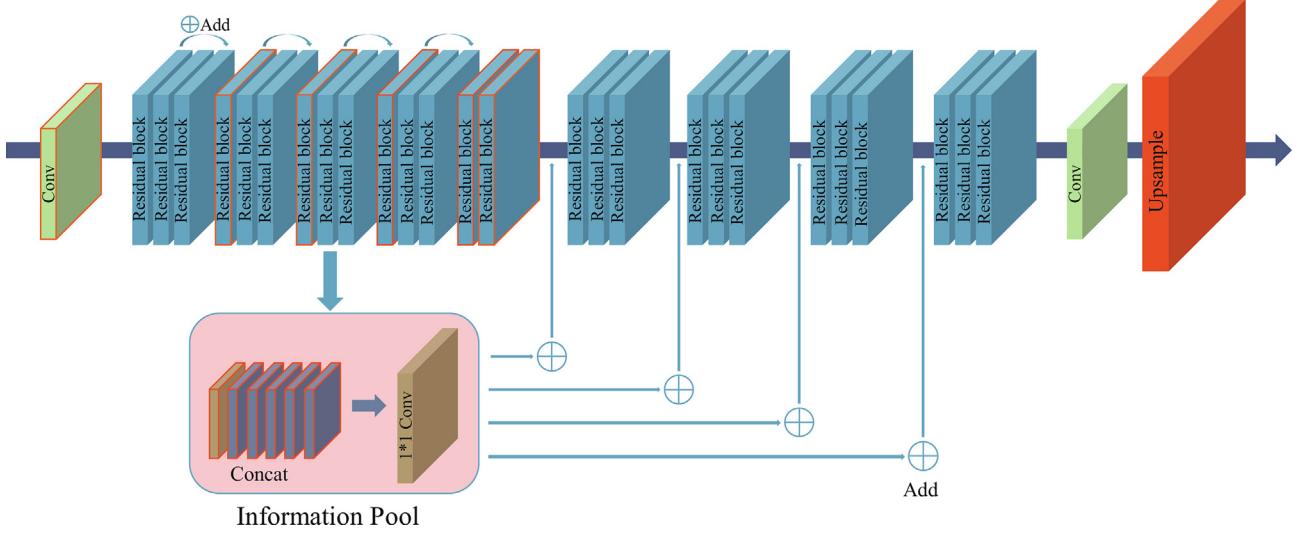


Fig. 5. The s-LWSR architecture [52].

Stage I: JPG Recovering (R). For low-quality JPG images, the compression operation discards a lot of useful details. As a result, how to rebuild details according to the original LR image plays a key role in this stage. To this end, we design a specialized network to learn the mapping between LR(C-JPG) and LR(normal) (All details are shown in Fig. 3).

Since the recovering task shares a similar goal with SR task, which is to restore accurate information in the pixel level, we employ an effective SR model, s-LWSR [52], to address the details recovering issue. The reason for introducing the SR model depends on the fact that all layers of the SR model are good at learning the representation of features in the pixel level. This is in accordance with the R stage which aims to recover information of all pixels. Learning from the stage I in Fig. 4, LR(C-JPG) images firstly go through the recovering model to accomplish the function of lossy details' reconstructing. In detail, a 3×3 convolutional layer transfers the basic RGB layers into fixed numbers. The main part of the recovering stage shares the same structure with our SR generation: stacking 26 residual blocks with additional information pool to intensively extract information. Experiments on s-LWSR have proven the powerful learning ability of this architecture. Finally, a convolution layer inversely realizes the transformation between the middle layers and the final three RGB layers.

Eventually, the recovered LR image for C-JPG achieves visual satisfactory in a certain degree, and the artificial trace appears smooth, which greatly reduces the pixel inconsistency. Furthermore, SR generation will benefit from these better quality LR inputs deriving from C-JPG in stage I.

Stage II: SR Generation (G). Because SR generation acts as a complementary task and is not the main contribution in our paper, we just bring in a state-of-the-art SR method called s-LWSR [52] in our model. To be specified, the most contribution of our research is the introduction of an application-dependent SR issue: how to get practicable SR images from C-JPG inputs.

As is mentioned, the s-LWSR leverages the combination of multi-level information from the front half of the model. More details are shown in Fig. 5. Specifically, the chosen layers are stacked to form the information pool, which transfers the combination of abundant low-level information to high-level layers through a series of activation operations.

Benefit from the R stage, the intermediate results greatly get rid of the artificial noise brought by the JPEG compression process. During the SR generation stage, intermediate images go through a similar model as the recovering stage. However, a specified

upsampling block, to ideally scale up the image is applied in the end. Normally, $2 \times$, $3 \times$, $4 \times$, and $8 \times$ are the most common SR tasks. Some models, such as MDSR [4], RCAN [7], can deal with all these scale-up tasks upon only one block with shared parameters. In this paper, like most DL-SR algorithms, we just study the single scale-up factor $4 \times$ SR problem to prove the effectiveness of the RGSR. As a result, the upsampling in our model contains two sub-pixel interpolation layers. In detail, each upsampling layer performs one $2 \times$ scale-up. Accordingly, experiments in the s-LWSR have proved that the mechanism can obtain SR generations with satisfactory visual quality.

3.4. Loss functions

Since the framework of our approach includes two typical tasks: JPG recovering and SR generation, there are three similarity loss functions as follows.

In stage I, each pixel in the C-JPG image corresponds to one supervised counterpart in the LR image. The goal is to recover the accurate value in pixels. As a result, we choose the $L1$ loss which is the most commonly used loss function in pixel-level tasks, like SR, denoising, deblurring, and style transfer. The function is formatted as:

$$\mathcal{L}_{L1}^1(I^{C-JPG}, I^{LR}) = \frac{1}{WH} \sum_{w=1}^W \sum_{h=1}^H \|I_{wh}^{LR} - G_{Rec}(I_{wh}^{C-JPG})\|, \quad (3)$$

where W and H refer to the width and height of the image respectively, and G_{Rec} is the transformation of stage I.

Based on the pre-processed C-JPG, we further scale the intermediate ($I^{LR_{rec}}$) up to the default size of the HR image through SR operation. Since our SR model pursues a better accurate SR generation, the loss function in stage II(G) inherits the same loss of former outstanding SR models. At first, DL-SR models use the $L2$ loss to calculate the difference. In the EDSR [4], the authors found that the $L1$ loss promotes better SR performance than former loss. Since then, almost all of the DL-SR models leverage the $L1$ loss to constrain the training. We set the loss as follows:

$$\mathcal{L}_{L1}^2(I^{LR_{rec}}, I^{HR}) = \frac{1}{S^2 WH} \sum_{w=1}^{sW} \sum_{h=1}^{sH} \|I_{sw \times sh}^{HR} - G_{SR}(I_{wh}^{LR_{rec}})\|, \quad (4)$$

where s is the scale factor.

Different from the normal SR task, C-JPG SR involves more difference between the input C-JPG and its final supervised HR image.

Inspired by the marvelous development of unsupervised style-to-style learning, such as CycleGAN [62], CinCGAN [13], and WESPE [63], we adopt the cycle loss to constrain the difference between the lossy input and the downsampled SR generations in our final loss. Taking CycleGAN as an example, images in different domains are transferred with inconsistent content. Cycle loss can keep the same content of the original image, while changing the style features, such as colors and textures. In this paper, SR generations are downsampled to input size to further compare with the corresponding non-JPG LR images. Here, we also use *L*₁ loss, and *Bic* refers to bicubic interpolation downsampling.

$$\mathcal{L}_{L1}^3(I^{SR}, I^{LR}) = \frac{1}{S^2WH} \sum_{w=1}^{SW} \sum_{h=1}^{SH} \|I_{wh}^{LR} - Bic(I_{sw \times sh}^{SR})\|_1. \quad (5)$$

How to balance the weights of all three losses shares a wide selection of possibilities. In fact, the loss severity during the JPEG compression is the key point. The more damages there are, the higher the weight of the stage R loss should be. To simulate the majority conditions, we choose equal importance as the default. As a result, the objective is the combination of the above three *L*₁ losses with equal weight:

$$L_{total} = \mathcal{L}_{L1}^1(I^{C-JPG}, I^{LR}) + \mathcal{L}_{L1}^2(I^{LR_{rec}}, I^{HR}) + \mathcal{L}_{L1}^3(I^{SR}, I^{LR}), \quad (6)$$

4. Experiments

In this section, we firstly describe the implementation details of our model. Secondly, we further analyze the effects of all functional

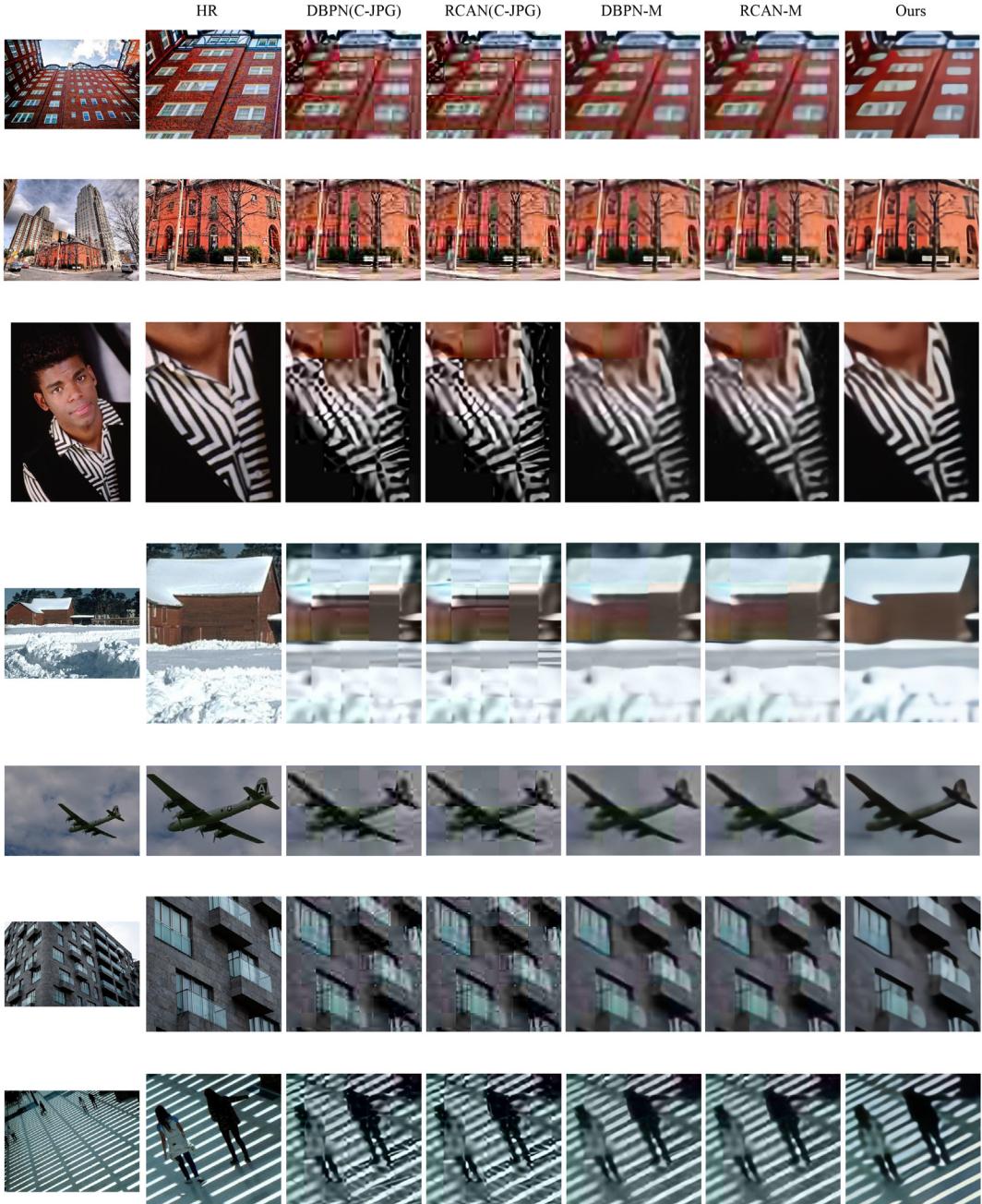


Fig. 6. Comparisons of our C-JPG SR model and other leading SR methods. Inputs from BSD100 [66] and Urban100 [41] are firstly processed into 20% quality JPG images (C-JPG) and Matlab pre-denoising images (M).

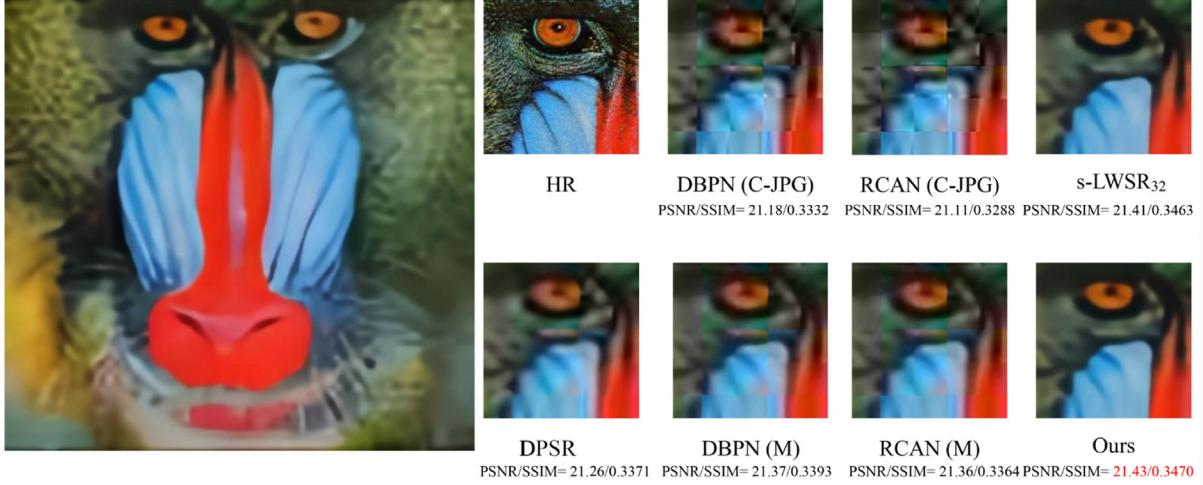


Fig. 7. The comparison of the RGSR and other SR methods. Inputs are firstly compressed to be 20% quality JPG images (C-JPG) and Matlab pre-denoising images (M). Both of PSNR and visual details indicate that our model achieves the best performance.

components through the extensive ablation study, in order to manifest the contribution of the proposed RGSR model. Thirdly, we compare our method with other leading SR models to validate its advantages. Finally, we exhibit the applications of our RGSR in reality conditions.

4.1. Implementation details

4.1.1. Dataset and pre-processing models

As mentioned in the previous part, our training process is based on three inputs: compressed LR (C-JPG), LR, and HR images. To conduct a fair comparison, we choose the DIV2K [8] as the training dataset, following the most leading SR models. There are 800 HR-LR training image pairs and 100 validation LR images in DIV2K. In order to obtain the C-JPG LR images, all LR images in the DIV2K are firstly restored with a specified quality setting by using the Pillow package in Python3.7. We also pre-process C-JPG images with the default deblurring model in the Matlab2018b. They are used to demonstrate the advantage of our recovering sub-model. Both of the Python and Matlab codes are available on <https://github.com/Sudo-Biao/RGSR>. We also provide the C-JPG DIV2K data link on the website.

4.1.2. Training settings

During training, images in C-JPG and LR dataset are randomly cropped into the size of 48×48 , while the size of cropped HR images is 192×192 . All LR images (C-JPG LR and LR) are processed with the data augmentation strategy which randomly rotates images by $90^\circ, 180^\circ, 270^\circ$, and flips horizontally. Considering the computation and time consumption, we set the batch size to be 16. Testing experiments are done on most popular SR datasets: Set5 [64], Set14 [65], BSD100 [66], and Urban100 [41]. Moreover, in order to prove the practical effectiveness, we perform our approach on the downloaded JPG images from the Internet. For both of the two stages, we apply the *de facto standard optimizer Adam* [67], where $\beta_1 = 0.9$ and $\beta_2 = 0.999$. The learning rate is initialized as 1×10^{-4} and half decreased every 2×10^2 iterations in back-propagation. Our model is trained 2×10^3 times until reaching its convergence. We implement our model RGSR by PyTorch with a Titan 2080Ti GPU.

4.1.3. Evaluation metrics

For accuracy-related SR models, most models evaluate the performance by the standards: PSNR and SSIM [68]. Recently, some SR

models aim to obtain SR images with realistic visual feelings, like SRGAN [20]. The standard for these methods is a non-reference assessment strategy, such as NIQE [69] and SSEQ [70]. In this paper, the RSGR pursues better accurate SR generations. As a result, all SR images are compared by the PSNR and the SSIM. Aligning with former SR algorithms, we calculate these two metrics on the Y channel (luminance) of the transformed YCbCr space instead of the original RGB channels.

4.2. Ablation study

To clarify the effectiveness of the recovering stage in our model, we apply the ablation study. To show the performance of our RGSR, several C-JPG inputs with various contents are tested. Then, two settings are employed to explore this ablation study: Firstly, the recovering stage is removed from the model. Secondly, we use the built-in denoising function of the Matlab2018b to replace the recovering stage.

4.2.1. Basic results

We evaluate the RGSR for $4\times$ enlarging task. To observe the performance under different conditions, we choose images with various contents: images of BSD100 [66], images of Urban100 [41], Images of Set14 [65], images of practical application, and random images from the Internet. The results are shown in Figs. 6, 7, 8.

It can be learned from these figures that our RGSR successfully removes the artificial trace while other models enlarge these unpleasant details. For example, in the third pic in Fig. 6, the black and white striped shirt appears to be smooth. To prove the accurate advantage of our RGSR, the SR images of several methods, such as DBPN [5], DPSR [10], RCAN [7], and our RGSR, are evaluated by PSNR and SSIM. All indexes confirm that our RGSR achieves the best performance. This advantage comes from the effect of the recovering stage. The influence of JPEG compression is restored by the recovering stage. Though it is not as accurate as of the HR image, it is satisfied for practical applications, such as printing.

4.2.2. Direct training

The recovering stage acts an important functional role in our model. In this part, we remove it from the model. As a result, C-JPG inputs directly go through stage II: SR generation. Correspondingly, we just train the model with \mathcal{L}_{L1}^2 loss. The remaining architecture is the same. As shown in Fig. 7, in the s-LWSR (training), many undesired artifacts are displayed, compared to our



Fig. 8. More comparisons of our RGSR and other leading SR methods. Inputs chosen from the Internet are firstly processing into 20% quality JPG images (C-JPG) and Matlab pre-denosing images (M). Our model achieves the best performance.

two-stage C-JPG model. Also, comparing on all four testing datasets is shown in Table 1. Either the PSNR or the SSIM decreases greatly. In fact, a single model is insufficient to resolve SR and recovering the C-JPG image simultaneously. The huge difference in supervised information leads to large variance therein the representations in the intermediate layers, which is crucial to the ideal details serving for the final SR operation.

4.2.3. Pre-denoising

To further analyze the learning ability of our recovering stage, we replace it by the built-in denoising code of the Matlab2018b. This setting comes from the fact that a lot of digital companies apply the Matlab to handle projects for its stability and strong computing power. Thanks to this mechanism, the C-JPG images

are firstly processed to a set of clean intermediate inputs: $I^{C-JPG} \rightarrow I^{inter}$. Then, we train the SR model with I^{inter} and I^{HR} pairs. The corresponding results are shown in Fig. 7. Although it clearly illustrates that pre-denoising operation can largely remove the artifacts, the setting is still obviously worse than our result. In our opinion, more details should be recovered instead of merely removing noise. Denoising only produces clearer C-JPG inputs without the learning course, while the recovering stage restores the accurate information to C-JPG images through learning.

4.3. Comparisons with leading methods

We provides quantitative evaluation of our model on four public benchmark datasets: Set5 [64], Set14 [65], BSD100 [66], and

Table 1

4 \times scaled performance comparison on the validation datasets for different methods (PSNR/SSIM). Scores in bold indicate the best performance. M refers to pre-processed inputs, and C-JPG means that the model is trained on 20% quality compressed JPG dataset.

Methods	Set5	Set14	BSD100	Urban100
DBPN	24.63/0.6437	23.33/0.5629	23.79/0.5540	21.26/0.5372
DBPN(M)	25.80/0.7135	24.17/0.6060	24.42/0.5735	22.08/0.5845
RCAN	24.45/0.6355	23.17/0.5559	23.65/0.5367	21.12/0.5302
RCAN(M)	25.76/0.7123	24.15/0.6051	24.40/0.5725	22.04/0.5836
s-LWSR ₃₂	24.73/0.6487	23.44/0.5669	23.86/0.5467	21.39/0.5424
s-LWSR ₃₂ (M)	25.82/0.7153	24.21/0.6087	24.43/0.5742	22.10/0.5860
s-LWSR ₃₂ (C-JPG)	26.34/0.7488	24.58/0.6275	24.63/0.5874	22.54/0.6186
Ours	26.37/0.7525	24.61/0.6282	24.64/0.5882	22.55/0.6208

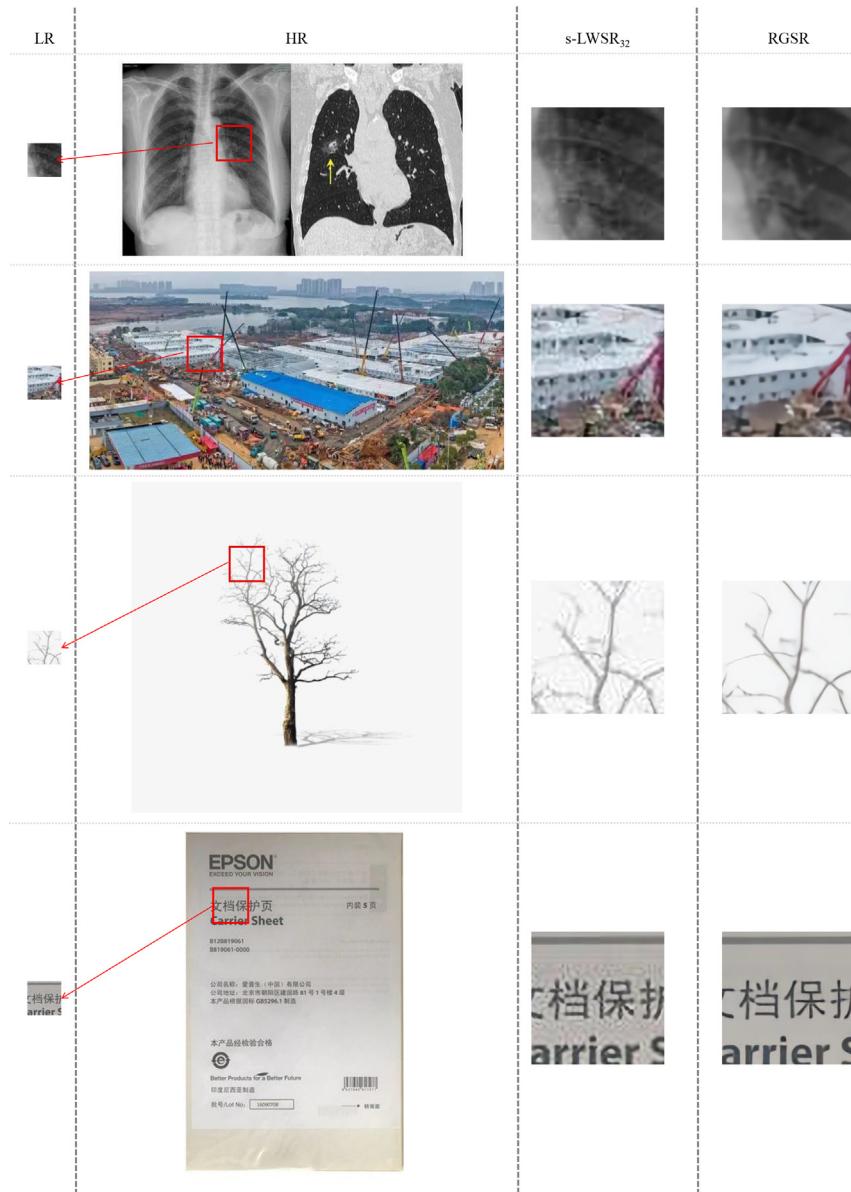


Fig. 9. More comparisons of our RGSR model and other SISR methods. The images are downloaded from the Internet without any change. The patch details shows that the RGSR generates smoother SR images.

Urban100 [41]. According to our goal, we firstly process images of the mentioned datasets to 20% quality JPG ones. Then, we compare our model with the state-of-the-art SR methods: DBPN [5], and RCAN [7]. Moreover, these methods are tested with pre-denoising inputs. We present all comparison results in Fig. 6 and Table 1. All index and visual details indicate that the proposed method (RGSR) surpasses all the former methods by a large margin. More slippy details and structured shapes exist in our SR generations. Our model successfully recovers detailed features and edges of I^{HR} , and exhibits satisfactory results compared to former SR models.

In detail, the pre-trained models enlarge the trace of 8×8 patch and result in poor performance. At the same time, the abandoned information gives fuzzy visual feelings, like the windows in the 2nd and 6th pics. The main reason for the phenomenon is that all parameters in the model represent the map between ideal LR images and HR images. The essential problem behind bad performance relates to the definition of SR issue. Our research comes

from the real problem when we try to print the SR images on cloth. The diffusion of pigments is worse than on papers. As a result, the trace of compression noise becomes extremely eye-catching. As a result, we study the issue in-depth and introduce the specialized SR issue: C-JPG SR.

To prove our opinion, we test the pre-trained models on the denoised C-JPG images. The setting of denoising is provided in 4.2.3. From Fig. 6 and Table 1, the effect of JPG compression relieved a lot. For example, the lines on the shirt in the third pic appear sharper; The zebra crossing in the last pic shows a clearer trace. The index also confirms the advantage that the PSNR of DBPN is 1.17dp higher on Set5.

In order to achieve better performance, we introduce two-step architecture to gradually solve the C-JPG issue. The RGSR is our continuing research on the SR issue. As a result, we utilize our former SR model: s-LWSR. Experiments exhibit the efficiency in the s-LWSR [52]. In fact, each part of the RGSR can be replaced by other SR models. The final performance of our RGSR surpasses other SR

models by a large margin (the PSRN 1.92 higher than RCAN on the set5). We also train the s-LWSR on the C-JPG dataset. The results in Table 1 confirm the specificity of the JPG SR problem. If we replace the sub-model of RGSR with other leading SR models, the generations will be as good as ours.

4.4. Applications in reality

Back to the starting point of the JPG SR problem, we apply the RGSR on several different images. To evaluate our model, images with various contents are downloaded from the Internet. To be specified, there is no supervised information for these images due to the absence of HR images. To better show the details of the improvement with RGSR in the generation of JPG-SR images, we compare the results by the exhibition of patches. It can be inferred from Figs. 8 and 9 that the RGSR greatly recover the missing details of C-JPG images. In Fig. 8, inputs are firstly compressed to 20% quality C-JPG images. Since the RGSR is trained on the 20% C-JPG dataset, it achieves outstanding performance than other SR models. To further prove the advantage of our RGSR, we apply the RGSR on four kinds of applications: CT image (ultrasound image), surveillance image, art design image, and phone photos (text). These scenarios greatly affect the application of the SR models. Results in Fig. 9 obviously show the great power of the RGSR. The clear windows in the second pic provide more information for the supervisor. In the 3rd and 4th pics, we can get the fact that the RGSR obtains pure pictures on art objects and text.

5. Conclusion

In this paper, we propose a SISR model (RGSR), for low-quality C-JPG images which is extensively used on the Internet. Based on our redefined C-JPG SR pipeline, two functional stages are integrated to fulfill the SR task on the C-JPG images. In addition, we employ cycle loss to guarantee the consistency between the above two stages. The intensive experiments demonstrate that our model can learn capable representations of LR inputs for the C-JPG SR task and outperform other cutting edges in SISR. To evaluate the advantage of the RGSR, we provide a lot of experiments on real conditions. Results confirm the value of our model. More exploration should be executed on other CV tasks with C-JPG images as the inputs in the future work.

CRediT authorship contribution statement

Biao Li: Supervision. **Yong Shi:** Methodology, Software, Data curation, Writing - original draft, Visualization. **Bo Wang:** Writing - review & editing, Visualization. **Zhiqian Qi:** Conceptualization, Writing - review & editing. **Jiabin Liu:** Data curation, 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 research is funded by the National Natural Science Foundation of China: No. 71932008, No. 71801232, No. 61702099, No.71501175, and No. 91546201, the Open Project of Key Laboratory of Big Data Mining and Knowledge Management, Chinese Academy of Sciences, and the Fundamental Research Funds for the Central Universities in UIBE (No. CXTD10-05).

References

- [1] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, A.L. Yuille, Deeplab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs, *IEEE Trans. Pattern Anal. Mach. Intell.* 40 (4) (2017) 834–848.
- [2] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778.
- [3] J. Kim, J. Kwon Lee, K. Mu 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.
- [4] 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.
- [5] 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.
- [6] 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.
- [7] 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.
- [8] 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.
- [9] K. Jiang, Z. Wang, P. Yi, G. Wang, K. Gu, J. Jiang, Atmfn: adaptive-threshold-based multi-model fusion network for compressed face hallucination, *IEEE Trans. Multimedia* (2019), 1–1.
- [10] K. Zhang, W. Zuo, L. Zhang, Deep plug-and-play super-resolution for arbitrary blur kernels, in: *IEEE Conference on Computer Vision and Pattern Recognition* (2019) 1671–1681.
- [11] J. Cai, H. Zeng, H. Yong, Z. Cao, L. Zhang, Toward real-world single image super-resolution: a new benchmark and a new model, *CoRR* abs/1904.00523.
- [12] K. Zhang, W. Zuo, L. Zhang, Learning a single convolutional super-resolution network for multiple degradations, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 3262–3271.
- [13] Y. Yuan, S. Liu, J. Zhang, Y. Zhang, C. Dong, L. Lin, Unsupervised image super-resolution using cycle-in-cycle generative adversarial networks, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2018, pp. 701–710.
- [14] J. Ma, X. Wang, L. Jiang, Image superresolution via dense discriminative network, *IEEE Trans. Ind. Electron.* 67 (7) (2019) 5687–5695.
- [15] P. Yi, Z. Wang, K. Jiang, Z. Shao, J. Ma, Multi-temporal ultra dense memory network for video super-resolution, *IEEE Trans. Circ. Syst. Video Technol.*
- [16] L. Zhou, Z. Wang, Y. Luo, Z. Xiong, Separability and compactness network for image recognition and superresolution, *IEEE Trans. Neural Networks Learn. Syst.* 30 (11) (2019) 3275–3286.
- [17] J. Jiang, Y. Yu, Z. Wang, S. Tang, R. Hu, J. Ma, Ensemble super-resolution with a reference dataset, *IEEE Trans. Cybern.* (2019) 1–15.
- [18] K. Zhang, W. Zuo, Y. Chen, D. Meng, L. Zhang, Beyond a gaussian denoiser: residual learning of deep cnn for image denoising, *IEEE Trans. Image Process.* 26 (7) (2017) 3142–3155.
- [19] Y. Zhang, K. Li, K. Li, B. Zhong, Y. Fu, Residual non-local attention networks for image restoration, *arXiv preprint arXiv:1903.10082*.
- [20] 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.
- [21] T. Köhler, M. Bätz, F. Naderi, A. Kaup, A.K. Maier, C. Riess, Benchmarking super-resolution algorithms on real data, *arXiv preprint arXiv:1709.04881*.
- [22] Z. Xu, Y. Shi, Exploring big data analysis: fundamental scientific problems, *Ann. Data Sci.*
- [23] T. Sirimongkolkaesem, R. Drikvandi, On regularisation methods for analysis of high dimensional data, *Ann. Data Sci.*
- [24] D.L. Olson, Y. Shi, *Introduction to Business Data Mining*, Irwin/mcgraw Hill.
- [25] Y. Shi, Y. Tian, G. Kou, Y. Peng, J. Li, *Optimization Based Data Mining: Theory and Applications*, first ed., Springer Publishing Company, Incorporated, 2011.
- [26] H.S. Dadi, G.K.M. Pillutla, M.L. Makkena, Face recognition and human tracking using gmm, hog and svm in surveillance videos, *Ann. Data Sci.* 5 (2) (2018) 157–179.
- [27] R. Keys, Cubic convolution interpolation for digital image processing, *IEEE Trans. Acoust. Speech Signal Process.* 29 (6) (1981) 1153–1160.
- [28] W.T. Freeman, T.R. Jones, E.C. Pasztor, Example-based super-resolution, *IEEE Comput. Graph. Appl.* 22 (2) (2002) 56–65.
- [29] 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.
- [30] C.-Y. Yang, J.-B. Huang, M.-H. Yang, Exploiting self-similarities for single frame super-resolution, in: *Asian Conference on Computer Vision*, Springer, 2010, pp. 497–510.
- [31] J. Yang, Z. Wang, Z. Lin, S. Cohen, T. Huang, Coupled dictionary training for image super-resolution, *IEEE Trans. Image Process.* 21 (8) (2012) 3467–3478.

- [32] R. Timofte, V. De Smet, L. Van Gool, Anchored neighborhood regression for fast example-based super-resolution, in: Proceedings of the IEEE International Conference on Computer Vision, 2013, pp. 1920–1927.
- [33] C.-Y. Yang, M.-H. Yang, Fast direct super-resolution by simple functions, in: Proceedings of the IEEE International Conference on Computer Vision, 2013, pp. 561–568.
- [34] H. Chang, D.-Y. Yeung, Y. Xiong, Super-resolution through neighbor embedding, in: Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004, vol. 1, IEEE, 2004, pp. I–I.
- [35] C. Dong, C.C. Loy, K. He, X. Tang, Learning a deep convolutional network for image super-resolution, in: European Conference on Computer Vision, Springer, 2014, pp. 184–199.
- [36] M. Ebrahimi, E.R. Vrcsay, Solving the inverse problem of image zooming using self-examples, in: International Conference Image Analysis and Recognition, Springer, 2007, pp. 117–130.
- [37] G. Freedman, R. Fattal, Image and video upscaling from local self-examples, ACM Trans. Graph. (TOG) 30 (2) (2011) 12.
- [38] D.G.S.B.M. Irani, Super-resolution from a single image, in: Proceedings of the IEEE International Conference on Computer Vision, Kyoto, Japan, 2009, pp. 349–356.
- [39] T. Michaeli, M. Irani, Nonparametric blind super-resolution, in: Proceedings of the IEEE International Conference on Computer Vision, 2013, pp. 945–952.
- [40] J. Yang, Z. Lin, S. Cohen, Fast image super-resolution based on in-place example regression, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2013, pp. 1059–1066.
- [41] 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.
- [42] K. Jiang, Z. Wang, P. Yi, J. Jiang, J. Xiao, Y. Yao, Deep distillation recursive network for remote sensing imagery super-resolution, Remote Sens. 10 (11).
- [43] C. Dong, C.C. Loy, X. Tang, Accelerating the super-resolution convolutional neural network, in: European Conference on Computer Vision, Springer, 2016, pp. 391–407.
- [44] X. Wang, K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, Y. Qiao, C. Change Loy, Esgan: Enhanced super-resolution generative adversarial networks, in: Proceedings of the European Conference on Computer Vision (ECCV), 2018.
- [45] K. Jiang, Z. Wang, P. Yi, G. Wang, T. Lu, J. Jiang, Edge-enhanced gan for remote sensing image superresolution, IEEE Trans. Geosci. Remote Sens. 57 (8) (2019) 5799–5812.
- [46] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial nets, in: Advances in Neural Information Processing Systems, 2014, pp. 2672–2680.
- [47] J. Johnson, A. Alahi, L. Fei-Fei, Perceptual losses for real-time style transfer and super-resolution, in: European Conference on Computer Vision, Springer, 2016, pp. 694–711.
- [48] S. Ioffe, C. Szegedy, Batch normalization: accelerating deep network training by reducing internal covariate shift, arXiv preprint arXiv:1502.03167.
- [49] R. Timofte, E. Agustsson, L. Van Gool, M.-H. Yang, L. Zhang, Ntire 2017 challenge on single image super-resolution: methods and results, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2017, pp. 114–125.
- [50] G. Huang, Z. Liu, L. Van Der Maaten, K.Q. Weinberger, Densely connected convolutional networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 4700–4708.
- [51] 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.
- [52] B. Li, B. Wang, J. Liu, Z. Qi, Y. Shi, s-LWSR: Super Lightweight Super-Resolution Network, IEEE Transactions on Image Processing 29 (2020) 8368–8380.
- [53] X. Chu, B. Zhang, H. Ma, R. Xu, J. Li, Q. Li, Fast, accurate and lightweight super-resolution with neural architecture search, arXiv preprint arXiv:1901.07261.
- [54] A.G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, H. Adam, Mobilenets: efficient convolutional neural networks for mobile vision applications, arXiv preprint arXiv:1704.04861.
- [55] O. Ronneberger, P. Fischer, T. Brox, U-net: convolutional networks for biomedical image segmentation, in: International Conference on Medical Image Computing and Computer-assisted Intervention, Springer, 2015, pp. 234–241.
- [56] B. Zoph, Q.V. Le, Neural architecture search with reinforcement learning, arXiv preprint arXiv:1611.01578.
- [57] A. Shocher, N. Cohen, M. Irani, zero-shot super-resolution using deep internal learning, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 3118–3126.
- [58] J.-Y. Zhu, T. Park, P. Isola, A.A. Efros, Unpaired image-to-image translation using cycle-consistent adversarial networks, in: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 2223–2232.
- [59] Y. Shi, B. Li, B. Wang, Z. Qi, J. Liu, Unsupervised single-image super-resolution with multi-gram loss, Electronics 8 (8) (2019) 833.
- [60] X. Xu, Y. Ma, W. Sun, Towards real scene super-resolution with raw images, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019, pp. 1723–1731.
- [61] D.A. Huffman, A method for the construction of minimum-redundancy codes, Proc. IRE 40 (9) (1952) 1098–1101.
- [62] T. Kim, M. Cha, H. Kim, J.K. Lee, J. Kim, Learning to discover cross-domain relations with generative adversarial networks, in: Proceedings of the 34th International Conference on Machine Learning–Volume 70, JMLR. org, 2017, pp. 1857–1865.
- [63] A. Ignatov, N. Kobyshev, R. Timofte, K. Vanhoey, L. Van Gool, Wespe: weakly supervised photo enhancer for digital cameras, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2018, pp. 691–700.
- [64] M. Bevilacqua, A. Roumy, C. Guillemot, M.L. Alberi-Morel, Low-complexity Single-image Super-resolution based on Nonnegative Neighbor Embedding, BMVA press, 2012.
- [65] R. Zeyde, M. Elad, M. Protter, On single image scale-up using sparse-representations, in: International Conference on Curves and Surfaces, Springer, 2010, pp. 711–730.
- [66] D. Martin, 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 Eighth IEEE International Conference on Computer Vision. ICCV 2001, vol. 2, IEEE, 2001, pp. 416–423.
- [67] D.P. Kingma, J. Ba, Adam: a method for stochastic optimization, Comput. Sci.
- [68] Z. Wang, A.C. Bovik, H.R. Sheikh, E.P. Simoncelli, et al., Image quality assessment: from error visibility to structural similarity, IEEE Trans. Image Process. 13 (4) (2004) 600–612.
- [69] A. Mittal, I.E.E.E. Fellow, R. Soundararajan, A.C. Bovik, Making a ‘completely blind’ image quality analyzer, IEEE Signal Process. Lett. 20 (3) (2013) 209–212.
- [70] L. Liu, B. Liu, H. Huang, A.C. Bovik, No-reference image quality assessment based on spatial and spectral entropies, Signal Process. Image Commun. 29 (8) (2014) 856–863.



Biao Li received a M.B.A. in Management from Tianjin University, Tianjin, China, in 2017. He received a Ph.D. in Management Science and Engineering from the University of the Chinese Academy of Sciences in 2020 under the supervision of Prof. Yong Shi. His research interests include machine learning and image processing.



Yong Shi received a Ph.D. in Management Science and Computer Systems from the University of Kansas, Lawrence, KS, USA. He is currently a professor with the Chinese Academy of Sciences, Beijing, China, where he serves as the Director of the Research Center on Fictitious Economy and Data Science. He is also a professor and a Distinguished Chair of Information Technology with the College of Information Science and Technology, University of Nebraska-Omaha, Omaha, USA. His research interests include data mining, information overload, optimal system designs, multiple-criteria decision making, decision support systems, and information and telecommunications management. He is the Editor-in-Chief of the International Journal of Information Technology and Decision Making and Annals of Data Science.



Bo Wang received a Ph.D. in mathematics from the University of the Chinese Academy of Sciences, Beijing, in 2014. He is currently an associate professor with the School of Information and Technology and Management, University of International Business and Economics, Beijing. His principal research interests include statistical machine learning, optimization-based data mining, and graphic models.



Zhiqian Qi is an associate professor with the Research Center on Fictitious Economy and Data Science, Chinese Academy of Sciences, Beijing, China. His research interests include object detection, object tracking, change detection, super-resolution, and machine learning.



Jiabin Liu received a Ph.D. from the University of the Chinese Academy of Sciences in 2019 under the supervision of Prof. Yong Shi. His work lies in data mining, the applications of weak label learning and image super-resolution.