GLOBAL REASONING NETWORK FOR IMAGE SUPER-RESOLUTION

Anonymous authors
Paper under double-blind review

ABSTRACT

Recent image super-resolution(SR) studies leverage very deep convolutional neural networks and the rich hierarchical features they offered, which leads to better reconstruction performance than conventional methods. However, the small receptive fields in the up-sampling and reconstruction process of those models stop them to take full advantage of global contextual information. This causes problems for further performance improvement. In this paper, inspired by image reconstruction principles of human visual system, we propose an image super-resolution global reasoning network (SRGRN) to effectively learn the correlations between different regions of an image, through global reasoning. Specifically, we propose global reasoning up-sampling module (GRUM) and global reasoning reconstruction block (GRRB). They construct a graph model to perform relation reasoning on regions of low resolution (LR) images.They aim to reason the interactions between different regions in the up-sampling and reconstruction process and thus leverage more contextual information to generate accurate details. Our proposed SRGRN are more robust and can handle low resolution images that are corrupted by multiple types of degradation. Extensive experiments on different benchmark data-sets show that our model outperforms other state-of-the-art methods. Also our model is lightweight and consumes less computing power, which makes it very suitable for real life deployment.

1 INTRODUCTION

Image Super-Resolution (SR) aims to reconstruct an accurate high-resolution (HR) image given its low-resolution (LR) counterpart. It is a typical ill-posed problem, since the LR to HR mapping is highly uncertain. In order to solve this problem, a large number of methods have been proposed, including interpolation-based (Zhang & Wu., 2006), reconstruction-based(Zhang et al., 2012), and learning-based methods (Timofte et al., 2013; 2014; Peleg & Elad., 2014; Schulter et al., 2015; Dong et al.; Huang et al., 2015; Tai et al., 2017; Tong et al., 2017; Zhang et al., 2018a; Dong et al., 2016).

In recent years, deep learning based methods have achieved outstanding performance in super-resolution reconstruction. Some effective residual or dense blocks (Huang et al., 2017; Zhang et al., 2018b; Wang et al., 2018; Lim et al., 2017; Ledig et al., 2017; Ahn et al.; Li et al., 2018) have been proposed to make the network wider and deeper and achieved better results. However, they only pay close attention to improving the feature extraction module, ignoring that the upsampling process with smaller receptive fields does not make full use of those extracted features. Small convolution receptive field means that the upsampling process can only perform super-resolution reconstruction based on local feature relationships in LR. As we all know, different features interact with each other, and features which are in different regions have corresponding effects on upsampling and reconstruction of a certain region. That is to say that a lot of information is lost in the process of upsampling and reconstruction due to the limitation of the receptive field, although the network extracts a large number of hierarchical features which are from low frequency to high frequency.

Chariker et al. (2016; 2018) show that the brain generates the images we see based on a small amount of information observed by the human eye, ranther than acquiring the complete data from the point-by-point scan of the retina. This process of generating an image is similar to a SR process. According to their thought, we add global information in SR reconstruction and propose to use relational reasoning to implement the process that the human visual system reconstructs images with observed global information. In general, extracting global information requires a large receptive

field. A large convolution receptive field usually requires stacking a large number of convolutional layers, but this method does not work in the upsampling and reconstruction process. Because this will produce a huge number of parameters.

Based on the above analysis, we propose an image super-resolution global reasoning network (SRGRN) which introduces the global reasoning mechanism to the upsampling module and the reconstruction layer. The model can capture the relationship between disjoint features of the image with a small respective field, thereby fully exploits global information as a reference for upsampling and reconstruction. We mainly propose global reasoning upsampling module (GRUM) and global reasoning reconstruction block (GRRB) as the core structure of the network. GRUM and GRRB first convert the LR feature map into N nodes, each of which not only represents a feature region in the LR image, but also contains the influence of pixels in other regions on this feature. Then they learn the relationship between the nodes and fuse the information of each node in a global scope. After that, GRUM learns the relationship between the channels in each node and amplifies the number of channels for the upsampling process. And then they convert N nodes into pixels with global reasoning information. Finally, GRUM and GRRB complete the upsampling and reconstruction process respectively.

In general, our work mainly has the following three contributions:

- We propose an image super-resolution global reasoning network (SRGRN) which draws on the idea of image reconstruction principles of human visual system. We mainly focus on the upsampling module and the reconstruction module. The model reconstructs SR images based on relational reasoning in a global scope.

- We propose a global reasoning upsampling module (GRUM) and global reasoning reconstruction block (GRRB), which construct a graph model to implement the relational reasoning among the feature regions in an image via 1D and 2D convolution, and finally adds the information obtained by global reasoning to each pixel. It can provide more contextual information to help generate more accurate details.

- Our proposed GRUM and GRRB are lightweight, which makes it suitable for real life deployment. More importantly, GRUM and GRRB balance the number of parameters and the reconstruction performance well. They can be easily inserted into other models.

2 Related Works

Deep CNN for SR and upsampling methods. Deep learning has achieved excellent performance in image super-resolution tasks. For the first time, Dong et al. applied convolutional neural networks to image SR. After this, Kim et al. proposed VDSR (Kima et al., 2016) and DRCN (Kim et al., 2016) which introduced residual learning to make the network depth reach 20 layers achieved significant improvement. And then more and more researchers are starting to pay attention to the improvement of the network feature extraction part. Lim et al. (2017) proposed EDSR and MDSR, which introduce residual scaling and remove unnecessary modules from the residual block. Concerned that the previous models only adopt the feather of the last CNN, Zhang et al. (2018b) proposed residual dense network to make full use of hierarchical features from each Conv layer. The above and most of the subsequent networks implement the upsampling based on either transposed convolution (Zeiler et al., 2010; Zeiler & Fergus., 2014) or sub-pixel convolution (Shi et al., 2016). Although these models have achieved good results, there exists a problem that these upsampling methods have only a small receptive field. This means that upsampling can only take advantage of contextual information within a small area.

Recently, researchers propose some new super-resolution upsampling process. LapSRN (Lai et al., 2017) allows low-resolution images to be directly input into the network for step-by-step amplification. Haris et al. (2018) exploit iterative up-and-down sampling layers and propose DBPN. Li et al. (2019) further explore the application of feedback mechanism (weight sharing) in SR and propose the SRFBN. These models have achieved a better reconstruction performance. However, the Conv layers in these upsampling modules still have only a small receptive field.

Global reasoning machansim. Recently, graph-based deep learning methods have begun to be widely used to solve relation reasoning. Santoro et al. propose Relation Networks (RN) (Santoro

et al., 2017) to solve problems that depend on relational reasoning. Liu et al. (2018) propose SIN, which implement a object detection using a graph model for structure inference. Furthermore, Chen et al. (2018) model a Global Reasoning unit that consists of five convolutions for image classification, semantic segmentation and video action recognition task. Considering that the human visual system generates images based on the observed global information is also a reasoning process. Moreover, correlation between feature regions can be obtained through relational reasoning, which makes each pixel in the generated SR image jointly determined by the information in a global scope. Therefore, we propose a global reasoning network for SR. We will detail our SRGRN in next section.

3 PROPOSED MODEL

According to Chariker et al. (2016; 2018), there is only little information transmitted from the retina to the visual cortex, and then the brain will reconstruct the real-world images based on the information received. We regard it as a reasoning process in a global scope. For image SR, the upsampling module constructs SR images base on features in LR images, which is substantially similar to detecting the category of each pixel of a SR image and generating these pixels based on contextual information of corresponding LR image. Due to the limitation of the convolution receptive field, only a small amount of contextual information can be utilized to generate HR images in most other models. This leads that many details in the HR image are not fine. Similarly, the above problem also exists in the reconstruction process. To solve these problems, we simulate the reasoning process that exists in human visual system, and then propose SRGRN to make full use of the contextual information to recover accurate details, which is achieved by constructing graph model and reasoning the relationship between these regions in an image.

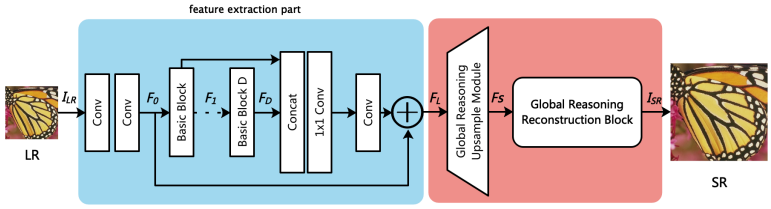

Figure 1: Super-resolution global reasoning network (SRGRN) architecture

3.1 ARCHITECTURE

As shown in Figure 1, our SRGRN includes feature extraction part, global reasoning upsample module(GRUM) and global reasoning reconstruction block(GRRB). Let's denote I_{LR} and I_{SR} as the input and output of SRGRN. The feature extraction part can use the relevant architecture of most other models. Here we introduce the feature extraction part of the RDN (Zhang et al., 2018b) as an example.

$$F_L = H_{FEX}(I_{LR}) \tag{1}$$

where $H_{FEX}(\cdot)$ denotes a series of operations of feature extraction part.

As with the previous work (Lim et al., 2017), the number of GRUM depends on the scaling factor, The GRUM receives F_L as input. F_S represents the output of the GRUM. GRUM can be expressed by the following mathematical formula:

$$F_S = H_{GRUM}(F_L) \tag{2}$$

where $H_{GRUM}(\cdot)$denotes a series of operations of GRUM. More details about GRUM will be given in Section 3.2.

We further conduct global reasoning reconstruction block(GRRB) to utilize the global contextual information to generate the output image. GRRB can be expressed by the following mathematical formula:

$$I_{SR} = H_{GRRB}(F_S) \tag{3}$$

where $H_{GRRB}(\cdot)$ denotes a series of operations of GRRB. More details about GRRB will be given in Section 3.3. After the above operations, we get the corresponding SR image.

3.2 GLOBAL REASONING UPSAMPLING MODULE

In this section, we present details about our proposed global reasoning upsample module(GRUM) in Figure 2. In order to help achieve relation reasoning, we map each image to a graph model

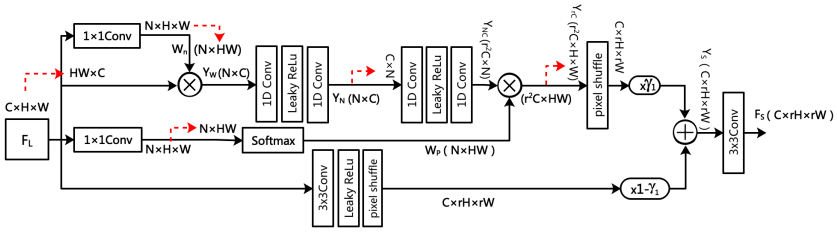

Figure 2: Global reasoning upsampling module (GRUM) architecture

$G = (V, E, u)$, where nodes in graph represent different regions in the image, E represents the relationship weight between the different regions, stored in the adjacency matrix of the graph, and u is the global information of the entire image.

We first need a function to construct *N* nodes in the oriented graph, each of which represents a region in the image. In GRUM, we obtain relationship weights between these pixels through a 2D 1×1 convolution, and then convert the input F_L into *N* nodes via element-wise product. The benefits of this approach are mainly reflected in the following aspects: (1) It can not only aggregates a feature region of the input F_L into a node, but also dig out the influence of other pixels in the image on this region. This is equivalent to adding global guidance of the image to each node. (2) Using convolution means that these relationship weights are trainable. This process can be expressed by the following mathematical formula:

$$W_n = V_d(Conv(F_L)) \tag{4}$$

$$Y_W = W_n \cdot (V_d(F_L))^T \tag{5}$$

where $F_L \in \mathbb{R}^{C \times H \times W}$ denotes the input tensor, $V_d(\cdot)$ denotes the operation of converting the vector shape from $N \times H \times W$ to $N \times HW$, $W_n \in \mathbb{R}^{N \times HW}$ denotes relationship weights and $Y_W \in \mathbb{R}^{N \times C}$ refers to *N* nodes with C channels.

After that, we use the 1D Conv - Leaky ReLu - 1D Conv (CLC) structure to implement reasoning and interaction between *N* nodes in the graph. The parameters in CLC refer to the adjacency matrix of the weighted oriented complete graph, which store the correlations between the nodes. CLC can learn and reason the complex nonlinear relationship between nodes better than only one 1D Conv. We use the following formula to describe the reasoning process between nodes:

$$Y_N = Conv(LRelu(Conv(Y_W))) \tag{6}$$

where $Conv(\cdot)$ and $LRelu(\cdot)$ denote 1D convolution along node-wise and Leaky ReLU (Maas et al., 2013) operation respectively.

And then we use the bottleneck to achieve channel amplification. The bottleneck receives $Y_N \in \mathbb{R}^{N \times C}$ as input and redistributes these channels by modeling the relationship between the channels of each node, amplifying the number of channel to $C \times r^2$, where r is the upscaling factor.

The first convolution in bottleneck makes channel C drop to $\bar{C} = C/\alpha$, where α represents reduction ratio. Then the second convolution makes the channel dimension $\bar{C}$ grow to $C \times r^2$. The bottleneck not only fits the complex relationships between channels better and redistributes channels more accurately, but also greatly reduces the number of parameters compared to the method of utilizing a single convolution. We use the following formula to describe channel amplification:

$$Y_{NC} = Conv(LRelu(Conv((Y_N)^T))) \tag{7}$$

where $Y_{NC} \in \mathbb{R}^{r^2 C \times N}$ refers to the output tensor.

In order to expand the resolution by pixelshuffle like ESPCN (Shi et al., 2016), we need to re-transform the *N* nodes ($r^2C \times N$) which have implemented the relational reasoning into a space whose shape is $C \times H \times W$. As above, we still learn a function to get a weight matrix whose shape is $N \times HW$ through a 1×1 2D convolution, and then normalize the weight matrix along the column with softmax. Finally, Y_{rC} and W_P can be obtained through:

$$W_P = Softmax(V_d(Conv(F_L))) \tag{8}$$

$$Y_{rC} = V_u(Y_{NC} W_P) \tag{9}$$

where $V_u(\cdot)$ denotes the operation of reshaping $N \times HW$ to $N \times H \times W$. $Y_{rC} \in \mathbb{R}^{r^2C \times H \times W}$ is a feature map where each pixel is associated with N nodes. $W_P \in \mathbb{R}^{N \times HW}$ is the normalized weight matrix. The value of these weights ranges from 0 to 1. This means that the reconstruction of each pixel is affected by N nodes to varying degrees. Each pixel in the feature map contains information which is generated by global reasoning.

After pixelshuffle, the output is multiplied by a parameter γ_1 and added to the upsampling result without global reasoning. The initial value of γ_1 is set to 0. As the global reasoning module trains, the network will gradually learn to assign values to the γ_1, thereby fully exploiting global reasoning. This process can be expressed by the following formula:

$$Y_S = H_{PS}(Y_{rC}) \cdot \gamma_1 + H_{UP}(F_L) \cdot (1 - \gamma_1) \tag{10}$$

where $H_{PS}(\cdot)$ denotes the operations of pixel shuffle and $H_{UP}(\cdot)$ denotes the operations of sub-pixel convolution.

Finally, F_S can be obtained by:

$$F_S = Conv(Y_S) \tag{11}$$

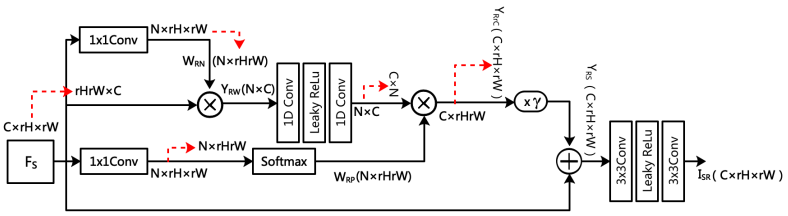

Figure 3: Global reasoning reconstruction block (GRRB) architecture

3.3 Global Reasoning Reconstruction Block

As shown in Figure 3, the specific details are similar to GRUM. We also construct a graph model for reconstruction block.

In GRRB, we first obtain the relationship weights $W_{RN} \in \mathbb{R}^{N \times rHrW}$ between pixels of F_S by 2D 1x1 convolution, and then aggregate the regions in F_S into N nodes by element-wise product. The output of this process $Y_{RW} \in \mathbb{R}^{N \times C}$ can be formulated as:

$$Y_{RW} = W_{RN} \cdot (V_d(F_S))^T = V_d(Conv(F_S)) \cdot (V_d(F_S))^T \tag{12}$$

After that, we use CLC to achieve the relationship reasoning between nodes. Then we exploit the weight matrix $W_{RP} \in \mathbb{R}^{N \times rHrW} = Softmax(V_d(Conv(F_S)))$ obtained by learning to redistribute the information of N nodes to the pixels. The output of this process $Y_{RrC} \in \mathbb{R}^{C \times rH \times rW}$ can be obtained by:

$$Y_{RrC} = V_u((F_{CLC}(Y_{RW}))^T \cdot W_{RP}) = V_u((F_{CLC}(Y_{RW}))^T \cdot Softmax(V_d(Conv(F_S)))) \tag{13}$$

where F_{CLC} refers to the operations of CLC. In addition, we apply the idea of residual connection in GRRB, which multiplys the information generated via global reasoning by a parameter γ and then add it to the input feature map. The output is given by:

$$Y_{RS} = \gamma \cdot Y_{RrC} + F_S \tag{14}$$

The initial value of γ is set to 0. As the training progresses, the network assigns more weight to γ. Finally, we input the feature map with global reasoning into the two Convs for reconstruction. We can get the final output through:

$$I_{SR} = H_{RL}(Y_{RS}) \tag{15}$$

where $H_{RL}(\cdot)$ denotes the operations of two Convs.

3.4 Implementation Details

In our proposed SRGRN, like the previous method (Lim et al., 2017), the number of GRUM depends on the scaling factor. For Conv layers with kernel size 3×3, we pad zeros to keep size fixed. We set the reduction ratio in bottleneck as α. The number of nodes in the graph model is set to N. We utilize Leaky ReLu (Maas et al., 2013) with a negative slope of 0.2 as non-linear activation function. The feature extraction part of the network are the same as the RDN (Zhang et al., 2018b) settings. The final Conv layer has 1 or 3 output channels, as we output gray or color HR images.

4 EXPERIMENTS

4.1 SETTINGS

Datasets and Metrics. We train all our models using 800 training images in the DIV2K (Agustsson & Timofte., 2017) dataset, which contains high-quality 2K images that can be used for image super-resolution task. And We use five standard benchmark datasets to evaluate PSNR and SSIM (Wang et al., 2004) metrics: Set5 (Bevilacqua et al., 2012), Set14 (Zeyde et al., 2010), B100 (Martin et al., 2001), Urban100 (Huang et al., 2015) and Manga109 (Y.Matsui et al., 2017). The SR results are evaluated on Y channel of transformed YCbCr space.

Degradation Models. In order to make a fair comparison with existing models, bicubic dowmsampling(denoted as **BI**) is regarded as a standard degradation model. We use it to generate LR images with scaling factor $\times 2$, $\times 3$, and $\times 4$ from ground truth HR images. To fully demonstrate the effectiveness of our model, we also use two other degradation models and conduct special experiments for them. Our second model, we defined it as **BD**, which blurs HR images with a Gaussian kernel of size 7×7 and a standard deviation of 1.6, and then downsamples the image with scaling factor $\times 3$. In addition to **BI** and **BD**, we also built the **DN** model, which first performs bicubic downsampling with scaling factor $\times 3$ and then adds Gaussian noise with a noise level of 30.

Training Setting. In each training batch, 16 LR RGB patches of size 48×48 are extracted as inputs. We perform data enhancement on the training images, which are randomly rotated by $90°$, $180°$, $270°$ and flipped horizontally. We use the Adam optimizer to update the parameters of the network with β_1 = 0.9, β_2 = 0.999, and $\epsilon = 10^{-8}$. For all layers in the network, the initial learning rate is set to 0.0001, and then the learning rate is halved every 200 epochs. We use the Pytorch framework to implement our model with Tesla P100.

4.2 ABLATION INVESTIGATION AND STUDY OF PARAMETERS

Global reasoning upsampling module. In order to verify the importance of the GRUM, we remove the GRUM from the network, leaving only the GRRB in the network for relation reasoning. As shown in Table 1, after removing GRUM, the performance of the network drops from 32.45 dB to 32.40 dB. When the Case Index is equal to 1, the corresponding model is the baseline model. We can observe that after GRUM is added to the baseline model, the network performance is improved from 32.31 dB to 32.42 dB. It can be seen that although our baseline model has achieved quite good results, GRUM can still improve the performance by relation reasoning in upsample module. This also indicates that relation reasoning can indeed result in better prformance. These comparisons fairly demonstrate the effectiveness of the GRUM for SR tasks.

Global reasoning reconstruction block. Then, we continue to study the effectiveness of GRRB for the network. After we add the GRRB to the baseline model, GRRB improves the performance of the model from 32.31 dB to 32.40 dB. Furthermore, the model with GRUM has achieved good performance. And it is difficult to obtain further improvements. But when we add the GRRB to it, the network performance shows a significant improvement, and the PSNR value on Set5 increases from 32.42 dB to 32.45 dB. These indicates that it is very essential for our network.

Basic parameters. Moreover, we also study the effects of two basic parameters N and α on the performance of the model. As shown in Table 1, we observe that larger N and smaller α would lead to higher performance. Considering that larger N and smaller α will also bring more computation, we set 10 and 8 as the value of N and α respectively.

Table 1: Ablation investigation of global reasoning upsampling method and global reasoning reconstruction block, study of N and α. We observe the best PSNR values on Set5 with scaling factor $\times 4$ in 200 epochs

Case Index	1	2	3	4	5	6	7	8
Global reasoning upsampling module	×	✓	×	✓	✓	✓	✓	✓
Global reasoning reconstruction block	×	×	✓	✓	✓	✓	✓	✓
Number of Node (N)	-	10	10	10	8	6	8	6
reduction ratio (α)	-	8	8	8	8	8	16	16
PSNR on Set5 ($4\times$)	32.31	32.42	32.40	32.45	32.43	32.42	32.42	32.40

4.3 NETWORK PARAMETERS

Several state-of-the-art methods are compared with our SRGRN in this section. We show comparisons about model size and performance in Figure 4. Although our SRGRN has less parameter

number than that of EDSR, MDSR and D-DBPN, our SRGRN and SRGRN+ achieve higher performance, having a better tradeoff between model size and performance. This demonstrates our method can well balance the number of parameters and the reconstruction performance.

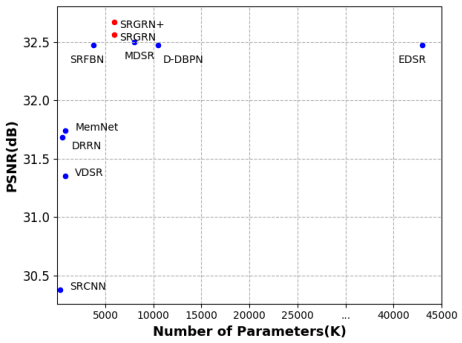

Figure 4: The results are evaluated on the Set5 dataset for 4 $\times$ SR. The proposed SRGRN strikes a balance between the number of parameters and the reconstruction effect.

Table 2: Average PSNR/SSIM for scale factors $\times 2$, $\times 3$ and $\times 4$ with BI degradation model. Best results are **highlighted**

Method	scale	Set5		Set14		BSDS100		Urban100		Manga109	
		PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Bicubic		33.66	0.9299	30.24	0.8688	29.56	0.8431	26.88	0.8403	30.80	0.9339
SRCNN		36.66	0.9542	32.45	0.9067	31.36	0.8879	29.50	0.8946	35.60	0.9663
DRRN		37.74	0.9591	33.23	0.9136	32.05	0.8973	31.23	0.9188	37.60	0.9736
LapSRN		37.52	0.9591	33.08	0.9130	31.08	0.8950	30.41	0.9101	37.27	0.9740
EDSR	$\times 2$	38.11	0.9602	33.92	0.9195	32.32	0.9013	32.93	0.9351	39.10	0.9773
RDN		32.24	0.9614	34.01	0.9212	32.34	0.9017	32.89	0.9353	39.18	0.9780
SRFBN		38.11	0.9609	33.82	0.9196	32.29	0.9010	32.62	0.9328	39.08	0.9779
SRGRN(ours)		38.25	0.9614	34.06	0.9214	32.37	0.9022	33.12	0.9367	39.31	0.9783
SRGRN+(ours)		**38.31**	**0.9616**	**34.15**	**0.9220**	**32.43**	**0.9027**	**33.32**	**0.9382**	**39.51**	**0.9786**
Bicubic		30.39	0.8682	27.55	0.7742	27.21	0.7385	24.46	0.7349	26.95	0.8556
SRCNN		32.75	0.9090	29.30	0.8215	28.41	0.7863	26.24	0.7989	30.48	0.9117
DRRN		34.03	0.9244	29.96	0.8349	28.95	0.8004	27.53	0.8378	32.42	0.9359
LapSRN		33.82	0.9227	29.87	0.8320	28.82	0.7980	27.07	0.8280	32.21	0.9350
EDSR	$\times 3$	34.65	0.9280	30.52	0.8462	29.25	0.8093	28.80	0.8653	34.17	0.9476
RDN		34.71	0.9296	30.57	0.8468	29.26	0.8093	28.80	0.8653	34.13	0.9484
SRFBN		34.70	0.9292	30.51	0.8461	29.24	0.8084	28.73	0.8641	34.18	0.9481
SRGRN(ours)		34.72	0.9297	30.60	0.8474	29.29	0.8102	28.95	0.8675	34.28	0.9491
SRGRN+(ours)		**34.79**	**0.9301**	**30.69**	**0.8487**	**29.36**	**0.8114**	**29.11**	**0.8705**	**34.58**	**0.9506**
Bicubic		28.42	0.8104	26.00	0.7027	25.96	0.6675	23.14	0.6577	24.89	0.7866
SRCNN		30.48	0.8628	27.50	0.7513	26.90	0.7101	24.52	0.7221	27.58	0.8555
DRRN		31.68	0.8888	28.21	0.7721	27.38	0.7284	25.44	0.7638	29.18	0.8914
LapSRN		31.54	0.8850	28.19	0.7720	27.32	0.7270	25.21	0.7560	29.09	0.8900
EDSR	$\times 4$	32.46	0.8968	28.80	0.7876	27.71	0.7420	26.64	0.8033	31.02	0.9148
RDN		32.47	0.8990	28.81	0.7871	27.72	0.7419	26.61	0.8028	31.00	0.9151
SRFBN		32.47	0.8983	28.81	0.7868	27.72	0.7409	26.60	0.8015	31.15	0.9160
SRGRN(ours)		32.56	0.8997	28.84	0.7880	27.75	0.7428	26.73	0.8053	31.13	0.9164
SRGRN+(ours)		**32.68**	**0.9009**	**28.95**	**0.7903**	**27.83**	**0.7445**	**26.94**	**0.8106**	**31.52**	**0.9197**

4.4 Results with BI Degradation Model

For BI degradation model, we compare our proposed SRGRN and SRGRN+ with other seven state-of-the-art image SR methods in quantitative terms. Following the previous works (Lim et al., 2017; Zhang et al., 2018b; Li et al., 2019), we also introduced a self-ensemble strategy to further improve the performance. We denote the self-ensemble method as SRGRN+.

A quantitative result for $\times 2$, $\times 3$, and $\times 4$ is shown in Table 2. We compare our models with other state-of-the-art methods on PSNR and SSIM. It can be seen that our proposed SRGRN outperforms other methods on all datasets without adding self-ensemble. After adopting self-ensemble, the performance further improves on the basis of SRGRN, and it achieved the best on all datasets. It is worth mentioning that SRFBN (Li et al., 2019) uses DIV2K+Flickr2K as their training set, which employs more training images than us. Previous research has come to a conclusion that more data in training set leads to a better result. However, their results are still not comparable to ours. Although RDN (Zhang et al., 2018b) is a state-of-the-art method, our SRGRN can achieve better performance in all datasets through relational reasoning in upsampling and reconstruction parts. The quantitative results indicate that our GRUM and GRRB play a vital role in improving network performance.

4.5 RESULTS WITH BD AND DN DEGRADATION MODELS

To show the robustness of the model, we also show the SR results with BD degradation model and further introduce DN degradation model. We compare 3 $\times$ SR results with other seven state-of-the-art image SR methods in quantitative terms. And We re-train SRCNN (Dong et al.) and VDSR (Kima et al., 2016) for **BD** and **DN** degradation model because of mismatched degradation model. For **BD** and **DN**, there is no doubt that reconstruction has become more difficult. As shown in Table 3 and Table 4, in the case of images with a lot of artifacts and noise, our SRGRN can get a excellent performance. This shows that SRGRN can effectively denoise and alleviate blurring artifacts. And when added to self-ensemble, SRGRN+ can achieve a better improvement.

Table 3: Average PSNR/SSIM for scale factors $\times 3$ with BD degradation model. Best results are **highlighted**

Method	scale	Set5		Set14		BSDS100		Urban100		Manga109	
		PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Bicubic		28.78	0.8308	26.38	0.7271	26.33	0.6918	23.52	0.6862	25.46	0.8149
SRCNN		32.05	0.8944	28.80	0.8074	28.13	0.7736	25.70	0.7770	29.47	0.8924
VDSR		33.25	0.9150	29.46	0.8244	28.57	0.7893	26.61	0.8136	31.06	0.9234
IRCNN_G		33.38	0.9182	29.63	0.8281	28.65	0.7922	26.77	0.8154	31.15	0.9245
IRCNN_C	$\times 3$	33.17	0.9157	29.55	0.8271	28.49	0.7886	26.47	0.8081	31.13	0.9236
SRMDNF		34.09	0.9242	30.11	0.8364	28.98	0.8009	27.50	0.8370	32.97	0.9391
RDN		34.57	0.9280	30.53	0.8447	29.23	0.8079	28.46	0.8581	33.97	0.9465
SRGRN(ours)		34.66	0.9286	30.60	0.8460	29.27	0.8090	28.71	0.8633	34.27	0.9480
SRGRN+(ours)		**34.78**	**0.9295**	**30.71**	**0.8476**	**29.35**	**0.8104**	**28.96**	**0.8673**	**34.65**	**0.9498**

Table 4: Average PSNR/SSIM for scale factors $\times 3$ with DN degradation model. Best results are **highlighted**

Method	scale	Set5		Set14		BSDS100		Urban100		Manga109	
		PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Bicubic		24.01	0.5369	22.87	0.4724	22.92	0.4449	21.63	0.4687	23.01	0.5381
SRCNN		25.01	0.6950	23.78	0.5898	23.76	0.5538	21.90	0.5737	23.75	0.7148
VDSR		25.20	0.7183	24.00	0.6112	24.00	0.5749	22.22	0.6096	24.20	0.7525
IRCNN_G		25.70	0.7379	24.45	0.6305	24.28	0.5900	22.90	0.6429	24.88	0.7765
IRCNN_C	$\times 3$	27.48	0.7925	25.92	0.6932	25.55	0.6481	23.93	23.93	26.07	0.8253
RDN		28.46	0.8151	26.60	0.7101	25.93	0.6573	24.92	0.7362	28.00	0.8590
SRFBN		28.53	0.8182	26.60	0.7144	25.95	0.6625	24.99	0.7424	28.02	0.8618
SRGRN(ours)		28.62	0.8195	26.66	0.7151	25.99	0.6632	25.14	0.7462	28.11	0.8642
SRGRN+(ours)		**28.67**	**0.8206**	**26.73**	**0.7168**	**26.07**	**0.6645**	**25.31**	**0.7496**	**28.31**	**0.8674**

4.6 SUPER-RESOLUTION ON REAL-WORLD IMAGES

To prove that our SRGRN can be widely used in the real world and performs robustly, we also conduct SR experiments on representative real-world images. We reconstruct some low resolution images in the real world that lack a lot of high frequency information. Moreover, in this case, the original HR images are not available and the degradation model is unknown either. Experiments show our SRGRN can recover finer and more faithful real-world images than other state-of-the-art methods under this bad condition. This further reflects the superiority of relation reasoning.

5 CONCLUSION

In this paper, inspired by the process of reconstructing images from the human visual system, we propose an super-resolution global reasoning network (SRGRN) for image SR, which aims at completing the reconstruction of SR images through global reasoning. We mainly propose global reasoning upsampling module (GRUM) and global reasoning reconstruction block (GRRB) as the core of the network. The GRUM can give the upsampling module the ability to perform relational reasoning in a global scope, which allows this process to overcome the limitations of the receptive field and recover more faithful details by analyzing more contextual information. The GRRB also enables the reconstruction block to make full use of the interaction between the regions and pixels to reconstruct SR images. We exploit SRGRN not only to handle low resolution images that are corrupted by three degradation model, but also to handle real-world images. Extensive benchmark evaluations demonstrate the importance of GRUM and GRRB. It also indicates that our SRGRN achieves superiority over state-of-the-art methods through global reasoning.

Acknowledgements. This research is supported by National Key R&D Program of China under grant 2018YFC0831503

REFERENCES

Eirikur Agustsson and Radu Timofte. Ntire 2017 challenge on single image super-resolution: Dataset and study. 2017.

Namhyuk Ahn, Byungkon Kang, and Kyung-Ah Sohn. Fast, accurate, and lightweight super-resolution with cascading residual network.

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

Logan Chariker, Robert Shapley, and Lai-Sang Young. Orientation selectivity from very sparse lgn inputs in a comprehensive model of macaque v1 cortex. 2016.

Logan Chariker, Robert Shapley, and Lai-Sang Young. Rhythm and synchrony in a cortical network model. 2018.

Yunpeng Chen, Marcus Rohrbach, Zhicheng Yan, Shuicheng Yan, Jiashi Feng, and Yannis Kalantidis. Graph-based global reasoning networks. 2018.

Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Learning a deep convolutional network for image super-resolution.

Chao Dong, Chen Change Loy, and Xiaoou Tang. Accelerating the super-resolution convolutional neural network. 2016.

Muhammad Haris, Greg Shakhnarovich, and Norimichi Ukita. Deep back projection networks for super-resolution. 2018.

Gao Huang, Zhuang Liu, Laurens van der Maaten, and Kilian Q. Weinberger. Densely connected convolutional networks. 2017.

Jia-Bin Huang, Abhishek Singh, and Narendra Ahuja. Single image super-resolution from transformed self-exemplars. 2015.

Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee. Deeply-recursive convolutional network for image super-resolution. 2016.

Jiwon Kima, Jung Kwon Lee, and Kyoung Mu Lee. Accurate image super-resolution using very deep convolutional networks. 2016.

Wei-Sheng Lai, Jia-Bin Huang, Narendra Ahuja, and Ming-Hsuan Yang. Deep laplacian pyramid networks for fast and accurate super-resolution. 2017.

Christian Ledig, Lucas Theis, Ferenc Huszar, Jose Caballero, Andrew Cunningham, Alejandro t' Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, and Wenzhe Shi. Photo-realistic single image super-resolution using a generative adversarial network. 2017.

Juncheng Li, Faming Fang, Kangfu Mei, and Guixu Zhang. Multi-scale residual network for image super-resolution. 2018.

Zhen Li, Jinglei Yang, Zheng Liu, Xiaomin Yang, Gwanggil Jeon, and Wei Wu. Feedback network for image super-resolution. 2019.

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

Yong Liu, Ruiping Wang, Shiguang Shan, and Xilin Chen. Structure inference net: Object detection using scene-level context and instance-level relationships. 2018.

Andrew L. Maas, Awni Y. Hannun, and Andrew Y. Ng. Rectifier nonlinearities improve neural network acoustic models. 2013.

David Martin, Charless Fowlkes, Doron Tal, and Jitendra Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. 2001.

Tomer Peleg and Michael Elad. A statistical prediction model based on sparse representations for single image super-resolution. 2014.

Adam Santoro, David Raposo, David G.T. Barrett, Mateusz Malinowski, Razvan Pascanu, Peter Battaglia, and Timothy Lillicrap. A simple neural network module for relational reasoning. 2017.

Samuel Schulter, Christian Leistner, and Horst Bischof. Fast and accurate image upscaling with super-resolution forests. 2015.

Wenzhe Shi, Jose Caballero, Ferenc Huszár, Johannes Totz, Andrew P. Aitken, Rob Bishop, Daniel Rueckert, and Zehan Wang. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. 2016.

Ying Tai, Jian Yang, and Xiaoming Liu. Image super-resolution via deep recursive residual network. 2017.

Radu Timofte, Vincent De Smet, and Luc Van Gool. Anchored neighborhood regression for fast example-based super-resolution. 2013.

Radu Timofte, Vincent De Smet, and Luc Van Gool. A+: Adjusted anchored neighborhood regression for fast super-resolution. 2014.

Tong Tong, Gen Li, Xiejie Liu, and Qinquan Gao. Image super-resolution using dense skip connections. 2017.

Xintao Wang, Ke Y, Shixiang W, Jinjin Gu, Yihao Liu, Chao Dong, Chen Change Loy, Yu Qiao, and Xiaoou Tang. Esrgan: Enhanced super-resolution generative adversarial networks. 2018.

Zhou Wang, Alan Conrad Bovik, Hamid Rahim Sheikh, and Eero P. Simoncelli. Image quality assessment: from error visibility to structural similarity. 2004.

Y.Matsui, K.Ito, Y.Aramaki, A.Fujimoto, T.Ogawa, T.Ya masaki, and K. Aizawa. Sketch-based manga retrieval using manga109 dataset. 2017.

Matthew D. Zeiler and Rob Fergus. Visualizing and understanding convolutional networks. 2014.

Matthew D. Zeiler, Dilip Krishnan, Graham W. Taylor, and Rob Fergus. Deconvolutional networks. 2010.

Roman Zeyde, Michael Elad, and Matan Protter. On single image scale-up using sparse-representations. 2010.

Kai Zhang, Wangmeng Zuo, and Lei Zhang. Learning a single convolutional super-resolution network for multiple degradations. 2018a.

Kaibing Zhang, Xinbo Gao, Dacheng Tao, and . Xuelong Li. Single image super-resolution with non-local means and steering kernel regression. 2012.

Lei Zhang and Xiaolin Wu. An edge-guided image interpolation algorithm via directional filtering and data fusion. 2006.

Yulun Zhang, Yapeng Tian, Yu Kong, Bineng Zhong, and Yun Fu. Residual dense network for image super-resolution. 2018b.

A APPENDIX

A.1 VISUAL RESULT

Visual comparison with BI degradtion model. As shown in Figure 5, we show a visual comparison on 4× SR. For image "img_078" from Urban100, we observe that most methods, even RDN and SRFBN, cannot recover these lattices and suffer from extremely severe blurring artifacts. Only our SRGRN can alleviate these blurring artifacts, recovers sharper and clearer edges and finer texture. For image "MukoukizuNoChonbo" from Manga109, There are heavy blurrings artifacts in all comparison methods, and the outline of some letters are broken. However, our proposed SRGRN can accurately recover these outlines, more faithful to the ground truth. The above comparison results are mainly due to the fact that SRGRN can enable upsampling and reconstruction modules to utilize more contextual information through relation reasoning.

Visual comparison with BD and DN degradtion model. In Figure 6, we show the comparison of SRGRN with other models in visual results. For image "img_014", we use bicubic upsampling to recover these images whose HR images are blurred with a Gaussian kernel before bicubic downsampling, then we obtain SR images with a lot of noticeable blurring artifacts. We have also observed that most methods, including RDN and SRFBN, do not clearly recover the lines around the window. Only our SRGRN can suppress blurring artifacts and recover these clear enough lines close to the ground truth by relation reasoning. For image "img_002", a large amount of noise corrupt the LR image and make it loss some detail. It can be seen that when using bicubic for upsampling, the obtained image not only has a large number of blurring artifacts but also a large amount of noise. However, we find that our SRGRN has great potential for removing noise efficiently and recover more detail. This fully demonstrates the effectiveness and robustness of our SRGRN for **BD** and **DN** degradation models.

Visual comparison on Real-World Images. In figure 7, the resolution of these images is so small that there is a lot of high frequency information missing from them. Moreover, in this case, the original HR images are not available and the degradation model is unknown either. For image "window"(with 200 × 160 pixels), only our SRGRN is able to recover sharper window edges and produce clearer SR image. For image "flower"(with 256 × 200 pixels), most other methods recover images whose upper left corner produces the edge of the pistil that looks unreal. And their edges of the petals in the whole image are very blurry. Our SRGRN can recovers sharper edges and finer details than other state-of-the-art methods. The above analysis indicate our model perform robustly unknown degradation models. This further reflects the superiority of relation reasoning.

Figure 5: **Visual comparison for** 4× **SR with BI model on "img_078" from Urban100 and "MukoukizuNoChonbo" from Manga109.**

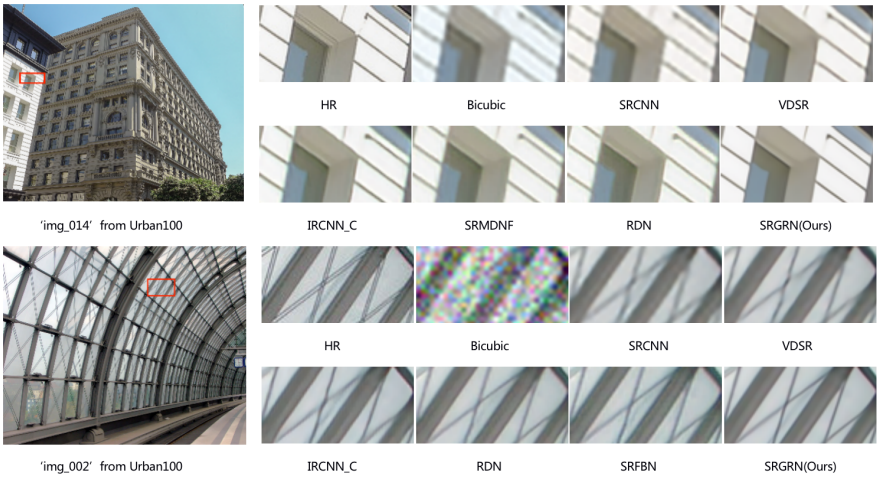

Figure 6: **Visual comparison for $3\times$ SR with BD model on "img_014" from Urban100 and for $3\times$ SR with DN model on "img_002" from Urban100**

Figure 7: **Visual comparison for 4 $\times$ SR on real-world images.** The two images show "flowers" and "window" respectively.