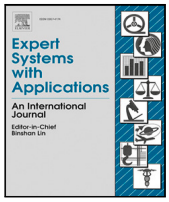
Expert Systems With Applications 250 (2024) 123954

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa



Lightweight blueprint residual network for single image super-resolution

Fangwei Hao [a,*], Jiesheng Wu [a], Weiyun Liang [a], Jing Xu [a], Ping Li [b]

[a] College of Artificial Intelligence, Nankai University, Tianjin, 300071, China
[b] Department of Computing, Hong Kong Polytechnic University, 999077, Hong Kong, China

ARTICLE INFO

Keywords:
Image super-resolution
Lightweight blueprint residual block
Effective feature fusion residual block
Lightweight blueprint residual network

ABSTRACT

The application of deep convolutional neural networks (CNNs) makes the lightweight single image super-resolution (SISR) task develop rapidly in recent years. However, existing lightweight SISR networks for the terminals with limited computing resources are still not efficient enough. In order to further alleviate the model complexity while keeping remarkable reconstruction performance, in this paper, we propose a lightweight blueprint residual network (LBRN), which is a powerful CNN-based SR model while simultaneously maintaining extremely lightweight. Specifically, we design a novel lightweight blueprint residual block (LBRB) to learn the high-frequency information efficiently for lightweight SISR. The LBRB with fewer parameters consists of more efficient convolution operations to extract high-frequency information which helps recover more visual details. In addition, we propose an effective feature fusion residual block (EFFRB) which is composed of three LBRBs with convolutional kernels of different sizes. The EFFRB, which is designed with effective multi-scale receptive fields, can availably extract and fuse multi-scale high-frequency information to obtain discriminative feature, and it further improves model performance. Extensive experiments show that our LBRN achieves superior reconstruction performance quantitatively and visually compared with other lightweight state-of-the-art SR methods.

1. Introduction

In recent years, deep learning has achieved great success in many computer vision tasks such as image classification (He et al., 2016; Cen et al., 2021; Liu et al., 2022; Zhai et al., 2022), object detection (Joseph et al., 2021; Li et al., 2021; Zheng et al., 2022; Gao et al., 2022; Lin et al., 2017), and semantic segmentation (Fan et al., 2021; Hoyer et al., 2022; Ma et al., 2022; Li et al., 2020). As a classical computer vision task, image super-resolution (SR) reconstruction, which aims at recovering a high-resolution (HR) image from the low-resolution (LR) counterpart, has been greatly improved since the application of the convolutional neural network (CNN). Such a task is an ill-posed problem that has attracted the attention of many researchers. Recently, many CNN-based networks (Dong et al., 2015; Kim et al., 2016a; Lim et al., 2017; Zhang et al., 2018; Hao et al., 2022; Zhang et al., 2020; Kim et al., 2016b; Tai et al., 2017; Ahn et al., 2018; Chu et al., 2021; Hui et al., 2018; Hui et al., 2019; Liu et al., 2020; Lan et al., 2020; Luo et al., 2022; Gao & Zhou, 2023) have emerged in single image super-resolution (SISR) for more accurately modeling the nonlinear mapping function from an LR to the HR. Dong et al. (2015) first designed a super-resolution convolutional neural network named SRCNN, which consisted of a three-layer CNN and obtained surprising reconstruction performance. For further improvements, A 20-layer network was designed by Kim et al. (2016a) and it achieved much higher effectiveness than SRCNN, which indicated that simply increasing the network depth can enhance the reconstruction performance. After the appearance of residual network (He et al., 2016), Lim et al. (2017) proposed a much deeper network termed as EDSR by introducing residual mechanism into one SR model. Even though its reconstruction performance was notably improved, a large number of model parameters were encountered. Later, more advanced networks were built, including the very deep residual channel attention network (RCAN) (Zhang et al., 2018) and the efficient residual attention network (ERAN) (Hao et al., 2022). By designing the residual in residual (RIR) structure and introducing the channel attention mechanism into the SR model, RCAN can bypass the massive low-frequency information in the LR image and pay attention to the high-frequency information for extracting more discriminative feature, and it obtained significant performance improvement with hundreds of convolution layers. ERAN further improved RCAN by proposing a channel hourglass residual structure (CHRS) and introducing the efficient channel attention (ECA) (Wang et al., 2020) mechanism into SISR.

* Corresponding author.
E-mail addresses: haofangwei@mail.nankai.edu.cn (F. Hao), jasonwu@mail.nankai.edu.cn (J. Wu), weiyunliang@mail.nankai.edu.cn (W. Liang), xujing@nankai.edu.cn (J. Xu), p.li@polyu.edu.hk (P. Li).

https://doi.org/10.1016/j.eswa.2024.123954
Received 19 June 2023; Received in revised form 2 March 2024; Accepted 8 April 2024
Available online 9 April 2024

Although deeper networks can obtain better performance, they suffer unfriendly computational burden when employed in portable applications. Therefore, it is important to design lightweight networks and improve the reconstruction efficiency for lightweight SISR task. In terms of lightweight image SR, various lightweight CNN-based SR methods (Kim et al., 2016b; Tai et al., 2017; Ahn et al., 2018; Chu et al., 2021; Hui et al., 2018; Hui et al., 2019; Liu et al., 2020; Lan et al., 2020; Luo et al., 2022; Gao & Zhou, 2023) have been proposed. Kim et al. (2016b) proposed a deeply-recursive convolutional network (DRCN), and Tai et al. (2017) proposed a deep recursive residual network (DRRN). They both embraced the recursive network, which shares network weights to reduce the number of model parameters. Although these two methods can reduce the model parameters, their performances degraded a lot. In order to make the model more lightweight, Ahn et al. (2018) proposed a cascading residual network named CARN-M for mobile devices, but it also leaded to a large drop in the peak signal-to-noise ratio (PSNR). Other lightweight SR endeavors (Hui et al., 2018; Hui et al., 2019; Liu et al., 2020) are about the information distillation SR networks, which mixed different features together to distill more useful information for the sequential blocks. The reconstruction performance and the model efficiency of these distillation SR networks were improved partly, which showed that the strategies of channel splitting and feature fusion were effective in lightweight SR. Nevertheless, they are still not very lightweight and lack powerful learning ability for discriminative features; that is, they are not efficient enough. Then, Lan et al. (2020) proposed a dense lightweight network named MADNet, which was developed to enhance the informative multi-scale feature representation ability and utilize the hierarchical features from original low-resolution images, yet it encountered a big drop in reconstruction performance. In addition, Luo et al. (2022) proposed the LatticeNet which consisted of the lattice blocks with butterfly-style topological structure to reduce model parameters; however, its extracted representation is not discriminative enough leading to the degradation of model performance. Later, Gao and Zhou (2023) proposed a very lightweight and efficient SISR network named VLESR to find a balance between the complexity and performance, but the reconstruction results was not satisfactory. Overall, existing lightweight SR methods are not efficient enough with significant drops in performance, the modules of these methods ignore the learning of high-frequency information which is crucial to recover more visual details, and they seldom take into account the effective fusion of multi-scale high-frequency feature to obtain discriminative feature.

To practically resolve these issues, we propose a lightweight blueprint residual network (LBRN), which is composed of lightweight blueprint residual blocks (LBRBs) and effective feature fusion residual blocks (EFFRBs), to improve the learning effectiveness and efficiency in lightweight SR. The LBRN can alleviate the model complexity while keeping remarkable reconstruction performance. Recently, a novel blueprint separation convolution (BSConv) (Haase & Amthor, 2020a) improved from depth-wise separable convolution (DSConv) (Howard et al., 2017) has shown its high learning efficiency in optimizing intra-kernel correlations for object recognition compared to standard convolution operation. The RIR structure in heavyweight SR method RCAN (Zhang et al., 2018) has been demonstrated to allow abundant low-frequency information to be bypassed and focus on learning high-frequency information, yet there is rare attention of designing a module to learning high-frequency information in lightweight SR. In view of this situation and inspired by these works (Zhang et al., 2018; Hao et al., 2022; Haase & Amthor, 2020a), we propose a novel lightweight blueprint residual block (LBRB), which incorporates the BSConv and the RIR structure, as the building block of our proposed LBRN. The LBRB can allow abundant low-frequency information to be bypassed and focus on extracting high-frequency information which helps recover more visual details. To make a further step, we propose an effective feature fusion residual block (EFFRB), which consists of three LBRBs with effective multi-scale receptive fields to extract and fuse multi-scale high-frequency information for obtaining discriminative feature, and it further improve model performance. Moreover, we designed different feature fusion modules as comparison and carry out extensive experiments to verify the superiority of the proposed EFFRB. Overall, our LBRN is more efficient compared to other lightweight state-of-the-art SR methods.

There are three main contributions offered in this paper:

(1) We propose a lightweight blueprint residual network (LBRN) to reconstruct a high-performance output from the corresponding LR for lightweight SR task. Our LBRN is more lightweight than most previous lightweight CNN-based networks while achieving superior SR performance.
(2) We propose a novel lightweight blueprint residual block (LBRB) by incorporating the BSConv and the RIR structure, and it is extremely lightweight and pays attention to the learning of high-frequency information which helps recover more visual details for lightweight SISR.
(3) We propose an effective feature fusion residual block (EFFRB) with effective multi-scale receptive fields. It consists of three LBRBs with convolutional kernels of different sizes to learn multi-scale high-frequency information, and the different high-frequency information is then fused for discriminative feature which further improves model performance.

The rest of the paper is organized as follows: Section 2 presents an overview of the related work. Section 3 details the proposed method. Section 4 shows the empirical research results. Section 5 shows the conclusion.

2. Related work

Recently, great progress has been made in image super-resolution (SR) based on deep learning. The pioneering deep SR work was proposed by Dong et al. (2015) and termed as SRCNN. It was a three-layer convolutional neural network (CNN) that can directly learn the mapping function from an LR to the corresponding HR. Compared to the early interpolation-based method (Zhang & Wu, 2006), the performance of SRCNN achieved a significant improvement quantitatively and visually due to the strong learning ability of CNN. Then, Kim et al. (2016a) pushed the network depth to 20 and proposed a very deep super-resolution (VDSR) network which obtained remarkable SR performance. In order to reduce the complexity of the model, Kim et al. (2016b) adopted the recursive network and proposed a deep recursive convolutional network (DRCN) for the SR task. Based on DRCN, Tai et al. (2017) proposed an improved one, i.e., a deep recursive residual network (DRRN), and it achieved better performance with fewer parameters. This indicated that the combination of the recursive mechanism and the residual module was relatively efficient. To address the problem of model efficiency, Ahn et al. (2018) proposed a laplacian pyramid super-resolution network (LapSRN) which had a unique network architecture. The LapSRN model was simplified and it achieved competitive performance compared to DRCN. For reducing the computational cost, Shi et al. (2016) designed an efficient sub-pixel convolution to upscale the resolution of features at the tail of the network, and they proposed an efficient sub-pixel convolutional neural network (ESPCN) to extract representative feature for SISR. Using the same post-upsampling strategy, a series of networks (Lim et al., 2017; Zhang et al., 2018; Hao et al., 2022; Zhang et al., 2020; Guo et al., 2020) were proposed to further improve the SR reconstruction. By removing unnecessary parts in conventional residual module (He et al., 2016), Lim et al. (2017) proposed an enhanced deep super-resolution network (EDSR), whose reconstruction performance reached a new level at the expense of significantly increasing the model parameters. Later, Zhang et al. (2020) proposed an residual dense network (RDN) by combining the residual learning and dense connection in the

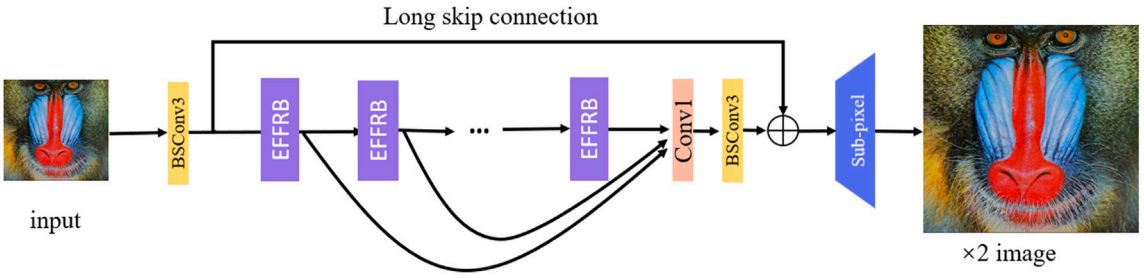


Fig. 1. Network architecture of our LBRN for ×2 SR.

residual dense block. They also proposed another network, i.e., a very deep residual attention network (RCAN) (Zhang et al., 2018), to exploit the application of the channel attention mechanism in SISR. Next, Guo et al. (2020) proposed a dual regression network named DRN, which introduced additional constraints, i.e., the closed-loop mappings so that the output LR images can enhance the performance of the SR model. Then, Hao et al. (2022) presented an efficient residual attention network (ERAN), which not only learned more representative features but also paid attention to network learning efficiency to some extent.

Although deep CNN-based SR methods have achieved great success, most of them are difficult in applying to mobile devices. To solve this issue, massive lightweight SR methods were proposed to simplify the SR model, and they drew the attention of many researchers, which promoted the development in this research field. Ahn et al. (2018) proposed a cascading residual network(CARN-M) for application in mobile devices. Hui et al. (2018) proposed an information distillation network (IDN) which embraced the strategy of channel splitting, and they forward presented the improved network, i.e., a lightweight information multi-distillation network (IMDN) (Hui et al., 2019). Another lightweight SR network based on information distillation strategy was the residual feature distillation network (RFDN) (Liu et al., 2020), which was a fast and lightweight image super-resolution method at that time. In addition, Lan et al. (2020) proposed a dense lightweight network named MADNet which can enhance representation ability of multi-scale feature and utilize the hierarchical features from original low-resolution images. Then, Luo et al. (2022) proposed the LatticeNet by designing a fancy lattice block and adopting the contrastive loss, and it achieved considerable performance. Next, Gao and Zhou (2023) proposed a very lightweight SISR network named VLESR which showed the exploration of model complexity and performance.

3. Proposed method

To pay more attention to learning the high-frequency information in lightweight SISR, we design a novel lightweight blueprint residual block (LBRB) as the building block. Furthermore, we propose an effective feature fusion residual block (EFFRB) to extract and fuse multi-scale high-frequency information to obtain discriminative feature. Finally, we propose a lightweight blueprint residual network named LBRN for more efficient super-resolution reconstruction.

3.1. Network architecture

As shown in Fig. 1, similar to RFDN (Liu et al., 2020), our LBRN is mainly made up of four parts: shallow feature extraction, effective feature fusion residual blocks (EFFRBs) for deep feature extraction, the feature fusion part and the reconstruction block. Let us suppose that the input and output of our network are denoted as I_{LR} and I_{SR}, respectively. Similar to Hui et al. (2018), Hui et al. (2019), Liu et al. (2020), given the input I_{LR}, its shallow feature F_0 is extracted by only one convolutional layer (Conv)

$$F_0 = H_s(I_{LR}) \tag{1}$$

where $H_s(\cdot)$ is the blueprint separation convolution (BSConv) operation. Then, the F_0 is gradually processed by M stacked EFFRBs, and each EFFRB refines its input feature and outputs the refined one. The output feature of the last EFFRB denotes the acquired deep feature at the stage of deep feature extraction, so we can formulate the process as

$$F_k = H_k(F_{k-1}), k = 1, \ldots, M \tag{2}$$

where $H_k(\cdot)$, F_{k-1} and F_k represent the kth EFFRB operation, its input feature and the output feature, respectively. After the F_0 is gradually refined by M EFFRBs, we use a 1×1 convolution layer to assemble all the intermediate feature, the assembled output is then processed by a 3×3 convolution layer for smoothing. the whole process can be formulated as follows

$$F_{assemble} = H_{assemble}(\text{Concat}(F_1, \ldots, F_M)) \tag{3}$$

where $\text{Concat}(\cdot)$ represents the concatenation operation along the channel dimension, $H_{assemble}(\cdot)$ denotes the operations of the 1×1 convolution layer and the followed 3×3 convolution layer, and $F_{assemble}$ is the assembled feature. At last, $F_{assemble}$ is processed by the reconstruction block at the tail, and the reconstruction process can be formulated as follows

$$F_{out} = R(F_{assemble} + F_0) \tag{4}$$

where $R(\cdot)$ is the reconstruction operation and F_{out} denotes the reconstruction output of the entire network. Specifically, the $R(\cdot)$ function consists of a 3×3 convolution layer and a sub-pixel convolution layer.

As for loss function, similar to RFDN (Liu et al., 2020), we also choose the L1 loss function for model optimization. Hence, the loss function of the LBRN is expressed as:

$$L(\Theta) = \frac{1}{N}\sum_{i=1}^{N} \| H_{LBRN}\left(I^i_{LR}\right) - I^i_{HR} \|_1 \tag{5}$$

Where $H_{LBRN}(\cdot)$ represents the function of the proposed LBRN, and Θ is the learnable parameters of our model. In order that the network training can converge quickly and effectively, we choose the adam (Kingma & Ba, 2014) optimization algorithm to optimize the entire network.

3.2. Lightweight blueprint residual block (LBRB)

The channel hourglass residual structure (CHRS) in ERAN (Hao et al., 2022) is a novel design which can bypass the abundant low-frequency information in the low-dimensional feature space, and can accelerate the flow of information across layers. It can make pixel-wise predictions without the upsampling or downsampling in the feature resolution. Blueprint separation convolution (BSConv) in Haase and Amthor (2020a) can optimize intra-kernel correlations and learn the features more efficiently compared with the standard convolution operation. Motivated by these methods, we design a novel lightweight blueprint residual block, i.e., the LBRB (see Fig. 2), which consists of P

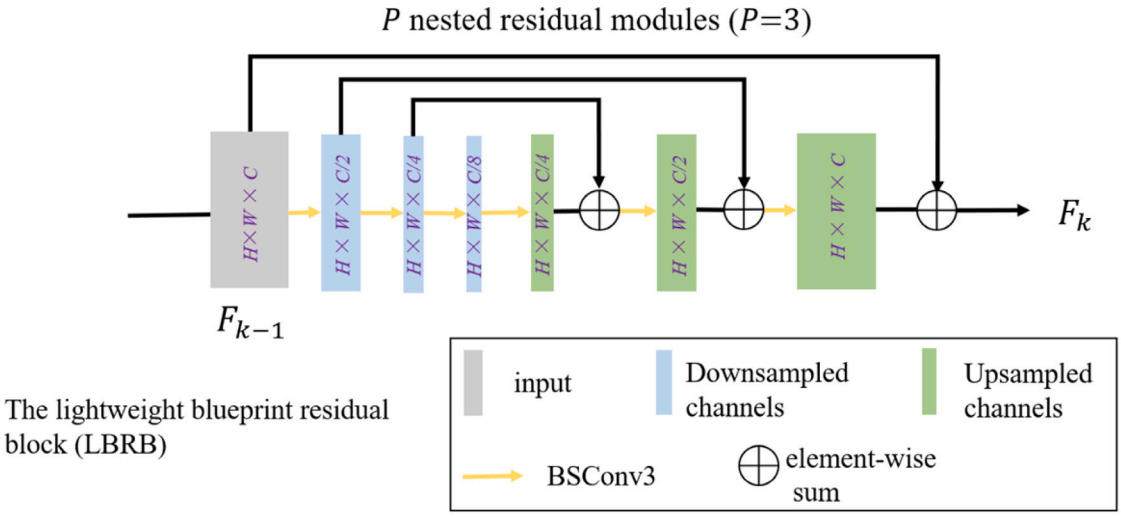


Fig. 2. The architecture of our lightweight blueprint residual block (LBRB), consisting of P=3 nested residuals.

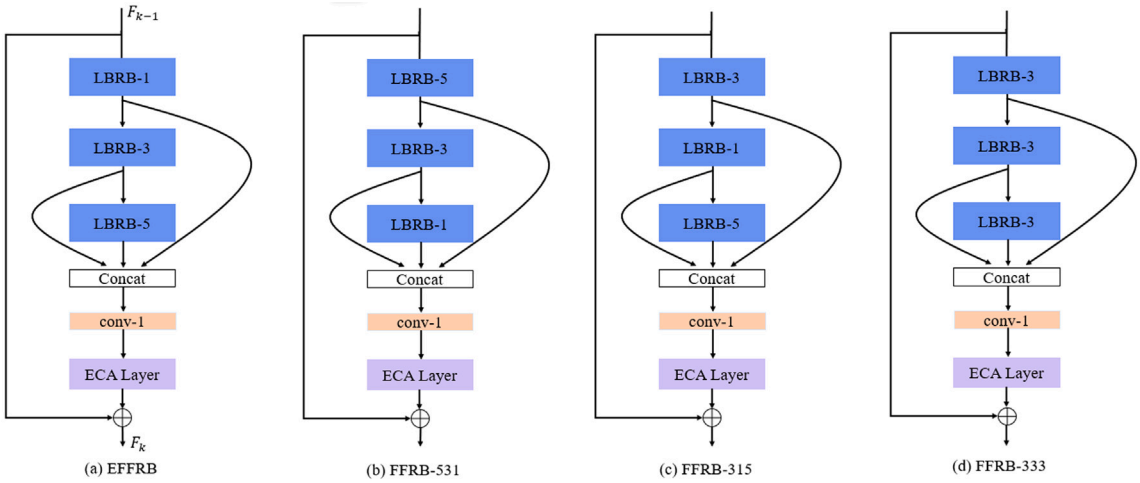


Fig. 3. The architectures of the EFFRB, FFRB-531, FFRB-315 and FFRB-333.

nested residuals to ease the learning of the high-frequency information for lightweight SISR.

We now show the LBRB in detail, let us denote F_{input} as its input feature maps with C channels and H×W size. In the downsampling part of LBRB, each layer performs channel downsampling on its input feature. On the contrary, in the upsampling part of a LBRB, each layer upsamples its corresponding input channels. In the inner operations of our LBRB, the resolution of all feature maps keeps the same, feature maps with the same number of channels are connected by skip connection, and all convolution layers are composed of blueprint separable convolutions (BSConvs) to formulate the intra-kernel correlations. As Fig. 2 shows, the LBRB with the RIR structure is a lightweight and efficient residual module for lightweight image SR, and it can focus on the learning of high-frequency information, while having fewer parameters than the CHRS under the same input size and output size. Different from RFDN (Liu et al., 2020) and ERAN (Hao et al., 2022), in our LBRB, all BSConv layers except the last are followed by the GReLU (Hendrycks & Gimpel, 2016) activation function.

3.3. Effective feature fusion residual block (EFFRB)

To explore the mechanism of effective multi-scale receptive fields among blocks for extracting discriminative feature, we propose an effective feature fusion residual block (EFFRB), and delicately design different combinations of multi-scale LBRBs as comparison, including different forward stacking order of three LBRBs and different sizes of convolution kernels.

we now clarify the details about the EFFRB (see Fig. 3). The EFFRB is made up of three LBRBs with convolution kernels of different sizes, i.e., 1×1, 3×3 and 5×5, which are denoted as LBRB-1, LBRB-3 and LBRB-5 respectively. Such design enables EFFRB to learn the multi-scale high-frequency information with effective multi-scale receptive fields, and we do not use larger convolution kernels due to considering the model complexity. The output high-frequency feature maps of three LBRBs are concatenated for feature fusion with one standard 1×1 convolution layer to obtain discriminative feature.

Specifically, an EFFRB contains three LBRBs, whose forward stacking order is LBRB-1, LBRB-3 and LBRB-5. By contrast, FFRB-531 consists of three LBRBs, whose forward stacking order is LBRB-5, LBRB-3 and LBRB-1; the forward stacking order of a FFRB-315 is LBRB-3, LBRB-1 and LBRB-5; and FFRB-333 is made up of three LBRB-3s without multi-scale operation.

As Fig. 3 shows, the receptive field of each LBRB in EFFRB is not the largest while larger convolution kernels are used to learn representation in a deeper block. Unlike this, the receptive field of each block in FFRB-531 reaches the maximum, and the deeper block is made up of smaller convolution kernels to extract feature maps. Then, the receptive fields of three blocks in a FFRB-315 are relatively irregular. In addition, each

Table 1
Investigations of LBRB, EFFRB, FFRB-531, FFRB-315 and FFRB-333 on the Set5 (×4), the best result is **highlighted**.

LBRB	×	✓	✓	✓	✓	✓
EFFRB	×	×	×	×	×	✓
FFRB-531	×	×	×	✓	×	×
FFRB-315	×	×	×	×	✓	×
FFRB-333	×	×	✓	×	×	×
PSNR (dB)	32.12	32.25	32.27	32.26	32.27	**32.33**

block of FFRB-333 consists of 3 × 3 convolution kernels without extracting multi-scale information. Experiments with these blocks, i.e., EFFRB, FFRB-531, FFRB-315 and FFRB-333 can demonstrate the most effective combination of multi-scale receptive fields and verify the ability of EFFRB for extracting discriminative feature.

4. Experiments

4.1. Settings

Following the previous works (Liu et al., 2020; Lan et al., 2020; Luo et al., 2022; Gao & Zhou, 2023), we use the DIV2K (Timofte et al., 2017) dataset which contains 800 high-quality RGB training images to train our networks from scratch. After finishing training, we test our models on five widely used benchmark datasets: SET5 (Bevilacqua et al., 2012), SET14 (Zeyde et al., 2012), BSDS100 (Arbelaez et al., 2010), URBAN100 (Huang et al., 2015) and MANGA109 (Matsui et al., 2017). After converting the SR results to YCbCr space, the peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM) (Wang et al., 2004) on the Y channel are adopted as evaluation metrics to show model performance quantitatively. Most of our experiments are based on the datasets generated by bicubic (BI) degradation model with scaling factors ×2, ×3 and ×4 for training and testing, some test experiments are conducted on Real-World Photos with compression artifacts (Lai et al., 2018; Zhang et al., 2018b). During training process, the ADAM (Kingma & Ba, 2014) optimizer with $\beta_1 = 0.9$, $\beta_2 = 0.99$ and $\epsilon = 10^{-8}$ is used to optimize our SR model. We conduct all experiments using Pytorch (Paszke et al., 2017) framework with an NVIDIA 3080Ti GPU. The learning rate is initialized as 10^{-3} and decays with a cosine annealing strategy.

4.2. Ablation studies

A series of experiments are conducted to analyze the efficiency of the proposed lightweight blueprint residual block (LBRB) and the effectiveness of the effective feature fusion residual block (EFFRB), compared with other three designed blocks including FFRB-531, FFRB-315 and FFRB-333.

First, we train the model without the LBRB, EFFRB, FFRB-531, FFRB-315 and FFRB-333 on the DIV2K dataset, and take the obtained performance value 32.12 dB PSNR as the baseline.

Next, we carry out verification experiments using the model with the LBRB only and the obtained corresponding result significantly increases by 0.13 dB reaching 32.25 dB PSNR, which demonstrates the powerful ability of the LBRB to improve the reconstruction performance due to focusing on the learning of the high-frequency information.

After adding LBRB to the baseline model, we further respectively add the EFFRB, FFRB-531, FFRB-315, and FFRB-333 to the model, and the corresponding experimental results are 32.33 dB PSNR, 32.26 dB PSNR, 32.27 dB PSNR, and 32.27 dB PSNR, while achieving improvements of 0.08 dB, 0.01 dB, 0.02 dB, and 0.02 dB respectively. These results demonstrate the effectiveness of multi-scale receptive fields for multi-scale high-frequency information, and they show that the EFFRB can improve SR reconstruction performance by generating discriminative feature.

The experimental results indicate that the model with LBRB and EFFRB achieves the best PSNR 32.33 dB, which demonstrates the effectiveness and superiority of the proposed LBRB and EFFRB. Specifically, our LBRB can significantly improve the reconstruction performance due to learning the high-frequency information; our EFFRB can extract discriminative features with more effective multi-scale receptive fields and outperform FFRB-531, FFRB-315 and FFRB-333. All experimental results verify the efficiency of our LBRN for lightweight SISR and they are shown in Table 1.

4.3. Comparisons with advanced methods

To further verify the effectiveness and efficiency of our LBRN, we compare our results quantitatively and visually with other state-of-the-art lightweight SR methods for upscaling factor ×2, ×3, and ×4, including SRCNN (Dong et al., 2015), FSRCNN (Dong et al., 2016), VDSR (Kim et al., 2016a), DRCN (Kim et al., 2016b), LapSRN (Lai et al., 2017), IDN (Hui et al., 2018), IMDN (Hui et al., 2019), PAN (Zhao et al., 2020), RFDN (Liu et al., 2020), MADNet (Lan et al., 2020), LatticeNet (Luo et al., 2022) and VLESR (Gao & Zhou, 2023).

PSNR/SSIM results. The quantitative evaluation results of lightweight ×2, ×3 and ×4 SR are showed in Table 2. For ×2 SR, most of the best quantitative performance are reached by our LBRN, namely, our LBRN achieves most of the highest PSNR and SSIM values on five benchmark datasets among these methods. For ×3 SR, our LBRN obtains the highest PSNR and SSIM values on all five benchmark datasets, and the same goes for ×4 SR. It can be seen that our LBRN can achieve superiority especially for a larger scaling factor (e.g., ×4) compared to other methods. Overall, all experimental results quantitatively show that our method can yield better reconstruction performance than previous lightweight state-of-the-art methods. Table 3 illustrates the comparison of model complexity and performance between the heavyweight state-of-the-art SR methods and our LBRN for ×4 SR on the benchmark datasets. As we can see, our model achieves competitive results compared to the heavy models while requiring at least ten times fewer parameters. Specifically, our LBRN achieved 27.60 dB which is only 0.11~0.23 dB lower than the heavy networks on the BSD100 dataset, while the difference becomes larger, reaching 0.13~0.41 dB, 0.18~0.36 dB and 0.44~0.86 dB on the Set5, Set14 and Urban100 datasets, respectively. These results suggest that our LBRN has an efficient trade-off between model complexity and performance.

Visual results. Fig. 4 shows the visual results of different methods on specific images of SET5 (Bevilacqua et al., 2012), BSDS100 (Arbelaez et al., 2010), URBAN100 (Huang et al., 2015) and MANGA109 (Matsui et al., 2017) datasets. For the image "butterfly.png", the original HR image contains a lot of sharp contour information. From the reconstruction results of these methods, it can be seen that the early bicubic algorithm (Zhang & Wu, 2006) has achieved the worst visual performance while other CNN-based methods have achieved much better results. For images "78004.png", "img_046.png" and "TaiyouNiSmash.png", the early bicubic method also produces widespread blurring and poor visual effect with aliasing artifacts and unstable trend, which indicates that the reconstruction quality of the early bicubic algorithm is undesirable. By contrast, the results of other methods (e.g., IMDN, RFDN and VLESR) show more contour information and clearer details. However, these CNN-based methods still present some distorted edges and blurring artifacts. In a word, the results show that the reconstruction algorithms based on the deep learning achieve better reconstruction performance than the early bicubic algorithm, which verify the powerful learning ability of CNN. Among these methods, our LBRN achieves the best visual perception and recovers sharper detail information of contour edges thanks to the more extracted high-frequency information and the obtained discriminative feature. These results visually demonstrate the effectiveness and superior reconstruction performance of our method.

Table 2
Quantitative results of state-of-the-art lightweight SR methods on benchmark datasets. The best and second-best results are **highlighted** and underlined, respectively.

Method	Scale	Params	Set5 PSNR/SSIM	Set14 PSNR/SSIM	BSD100 PSNR/SSIM	Urban100 PSNR/SSIM	Manga109 PSNR/SSIM
Bicubic		–	33.66/0.9299	30.24/0.8688	29.56/0.8431	26.88/0.8403	30.80/0.9339
SRCNN		8K	36.66/0.9542	32.45/0.9067	31.36/0.8879	29.50/0.8946	35.60/0.9663
FSRCNN		13K	37.00/0.9558	32.63/0.9088	31.53/0.8920	29.88/0.9020	36.67/0.9710
VDSR		666K	37.53/0.9587	33.03/0.9124	31.90/0.8960	30.76/0.9140	37.22/0.9750
DRCN		1774K	37.63/0.9588	33.04/0.9118	31.85/0.8942	30.75/0.9133	37.55/0.9732
LapSRN		251K	37.52/0.9591	32.99/0.9124	31.80/0.8952	30.41/0.9103	37.27/0.9740
IDN	×2	553K	37.83/0.9600	33.30/0.9148	32.08/0.8985	31.27/0.9196	38.01/0.9749
IMDN		694K	38.00/0.9605	33.63/0.9177	32.19/0.8996	32.17/0.9283	38.88/0.9774
PAN		261K	38.00/0.9605	33.59/0.9181	32.18/0.8997	32.01/0.9273	38.70/0.9773
RFDN		534K	38.05/0.9606	33.68/0.9184	32.16/0.8994	32.12/0.9278	38.88/0.9773
MADNet		878K	37.94/0.9604	33.46/0.9167	32.10/0.8988	31.74/0.9246	–/–
LatticeNet		756K	38.06/0.9607	**33.70/0.9187**	32.20/0.8999	32.25/0.9288	–/–
VLESR		311K	38.01/0.9605	33.58/0.9177	32.16/0.8993	32.14/0.9280	38.75/0.9770
LBRN (ours)		325K	**38.08/0.9608**	33.57/0.9173	**32.23/0.9005**	**32.35/0.9303**	**38.93/0.9777**
Bicubic		–	30.39/0.8682	27.55/0.7742	27.21/0.7385	24.46/0.7349	26.95/0.8556
SRCNN		8K	32.75/0.9090	29.30/0.8215	28.41/0.7863	26.24/0.7989	30.48/0.9117
FSRCNN		13K	33.18/0.9140	29.37/0.8240	28.53/0.7910	26.43/0.8080	31.10/0.9210
VDSR		666K	33.66/0.9213	29.77/0.8314	28.82/0.7976	27.14/0.8279	32.01/0.9340
DRCN		1774K	33.82/0.9226	29.76/0.8311	28.80/0.7963	27.15/0.8276	32.24/0.9343
LapSRN		502K	33.81/0.9220	29.79/0.8325	28.82/0.7980	27.07/0.8275	32.21/0.9350
IDN	×3	553K	34.11/0.9253	29.99/0.8354	28.95/0.8013	27.42/0.8359	32.71/0.9381
IMDN		703K	34.36/0.9270	30.32/0.8417	29.09/0.8046	28.17/0.8519	33.61/0.9445
PAN		261K	34.40/0.9271	30.36/0.8423	29.11/0.8050	28.11/0.8511	33.61/0.9448
RFDN		541K	34.41/0.9273	30.34/0.8420	29.09/0.8042	28.21/0.8525	33.67/0.9449
MADNet		930K	34.26/0.9262	30.29/0.8410	29.04/0.8033	27.91/0.8464	–/–
LatticeNet		765K	34.40/0.9272	30.32/0.8416	29.10/0.8049	28.19/0.8513	–/–
VLESR		319K	34.40/0.9272	30.34/0.8415	29.08/0.8043	28.16/0.8519	33.61/0.9445
LBRN (ours)		339K	**34.43/0.9276**	**30.39/0.8429**	**29.13/0.8059**	**28.29/0.8545**	**33.72/0.9455**
Bicubic		–	28.42/0.8104	26.00/0.7027	25.96/0.6675	23.14/0.6577	24.89/0.7866
SRCNN		8K	30.48/0.8626	27.50/0.7513	26.90/0.7101	24.52/0.7221	27.58/0.8555
FSRCNN		13K	30.72/0.8660	27.61/0.7550	26.98/0.7150	24.62/0.7280	27.90/0.8610
VDSR		666K	31.35/0.8838	28.01/0.7674	27.29/0.7251	25.18/0.7524	28.83/0.8870
DRCN		1774K	31.53/0.8854	28.02/0.7670	27.23/0.7233	25.14/0.7510	28.93/0.8854
LapSRN		502K	31.54/0.8852	28.09/0.7700	27.32/0.7275	25.21/0.7562	29.09/0.8900
IDN	×4	553K	31.82/0.8903	28.25/0.7730	27.41/0.7297	25.41/0.7632	29.41/0.8942
IMDN		715K	32.21/0.8948	28.58/0.7811	27.56/0.7353	26.04/0.7838	30.45/0.9075
PAN		272K	32.13/0.8948	28.61/0.7822	27.59/0.7363	26.11/0.7854	30.51/0.9095
RFDN		550K	32.24/0.8952	28.61/0.7819	27.57/0.7360	26.11/0.7858	30.58/0.9089
MADNet		1002K	32.11/0.8939	28.52/0.7799	27.52/0.7340	25.89/0.7782	–/–
LatticeNet		777K	32.18/0.8943	28.61/0.7812	27.57/0.7355	26.14/0.7844	–/–
VLESR		331K	32.17/0.8945	28.55/0.7802	27.55/0.7345	26.03/0.7830	30.48/0.9073
LBRN (ours)		345K	**32.33/0.8964**	**28.62/0.7826**	**27.60/0.7377**	**26.17/0.7882**	**30.60/0.9102**

Table 3
Quantitative comparison of our LBRN with the previous advanced heavy SR methods on the benchmark datasets for ×4 SR.

Method	Param	Set5	Set14	BSD100	Urban100
EDSR	43.1M	32.46 dB	28.80 dB	27.71 dB	26.64 dB
RDN	22.3M	32.47 dB	28.81 dB	27.72 dB	26.61 dB
RCAN	15.6M	32.63 dB	28.87 dB	27.77 dB	26.82 dB
DRN	9.8M	32.74 dB	28.98 dB	27.83 dB	27.03 dB
ERAN	8.02M	32.66 dB	28.92 dB	27.79 dB	26.86 dB
LBRN	0.35M	32.33 dB	28.62 dB	27.60 dB	26.17 dB

4.4. Experiments on real-world photos

In addition to the above experiments on synthesized LR images downsampled from HR images with BI degradation, we also do experiments on real LR images to demonstrate the effectiveness of proposed LBRN. Since there are none ground-truth HR images, we only provide the SR outputs for visual comparison. Figs. 5 and 6 show the SISR results on two real LR images "Chip" and "Historical-006", respectively. It can be seen that the proposed LBRN obtains sharper details and clearer images compared with the previous advanced methods including Bicubic (Zhang & Wu, 2006), SRCNN (Dong et al., 2015), IMDN (Hui et al., 2019), RFDN (Liu et al., 2020) and VLESR (Gao & Zhou, 2023). It is worth noting that the results of earlier methods, such as Bicubic and SRCNN, present a lot of compression artifacts while our LBRN can reconstruct sharper details with fewer artifacts, and can obtain more natural results due to the more extracted high-frequency information.

4.5. Model complexity analysis

In this section, we will show the detailed differences of different methods in model complexity and reconstruction performance (see Table 4 and Fig. 7) on Set5 dataset for ×4 SR. Model parameters and multi-adds quantitatively show the complexity of one SR model, and the evaluation metric PSNR quantitatively verifies the reconstruction quality. As Table 4 and Fig. 7 show, our LBRN can achieve the highest PSNR value with the second fewest model parameters among these methods including LapSRN (Lai et al., 2017), IMDN (Hui et al., 2019), RFDN (Liu et al., 2020), LatticeNet (Luo et al., 2022), and VLESR (Gao & Zhou, 2023). This demonstrates the effectiveness of our proposed modules and network. It is worth noting that our LBRN with 345K parameters is a little more complex than VLESR with minimum parameters 331K, but outperforms VLESR by a large gap 0.16 dB on Set5 dataset for ×4 SR.

Another quantitative evaluation metric for model complexity is the multi-adds which is defined in Ahn et al. (2018) and it denotes the number of multiply accumulate operations. Following LatticeNet, the SR output size is assumed to 1280×720 to calculate model multi-adds. To get a more comprehensive understanding of the model complexity,

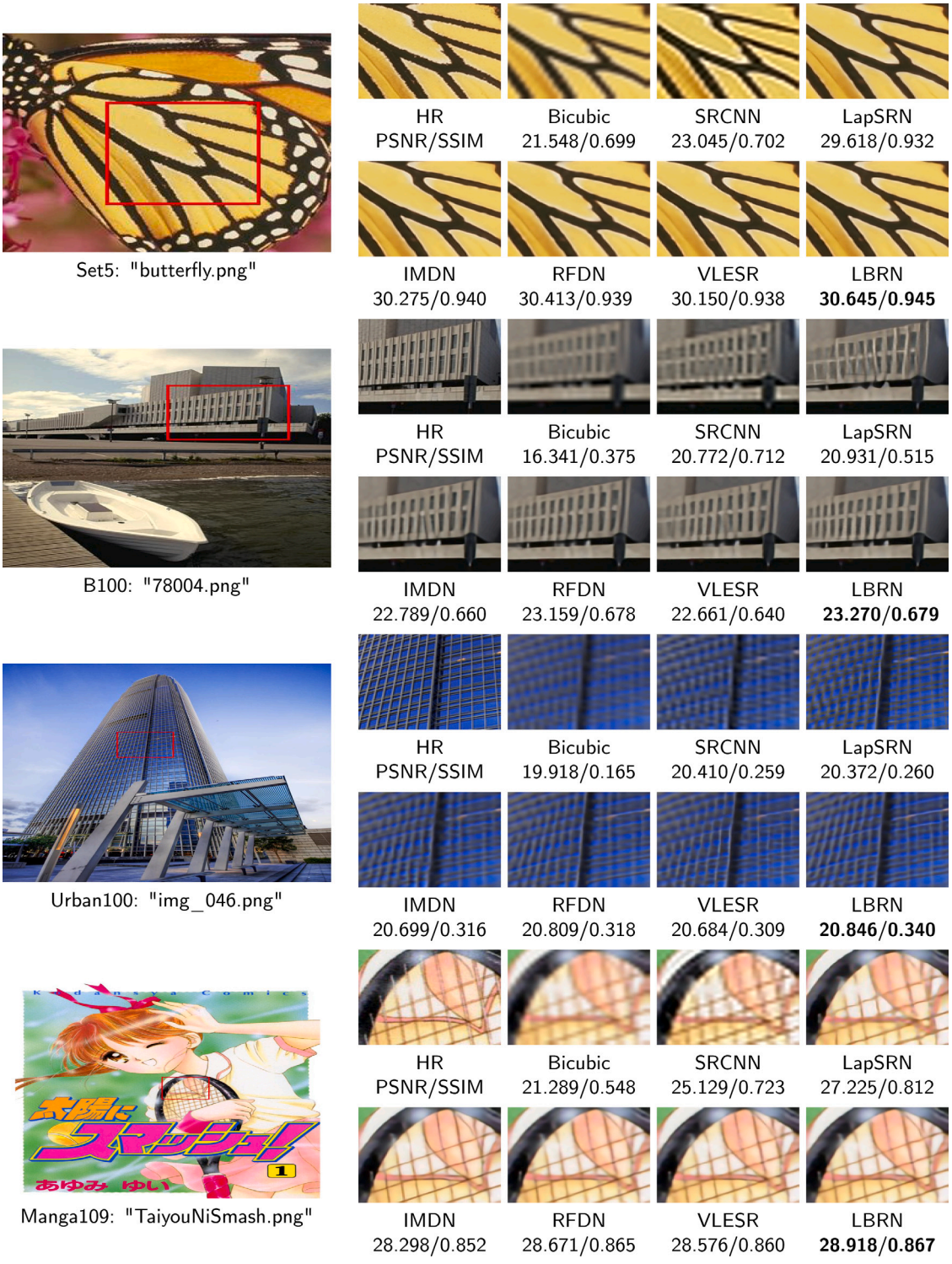
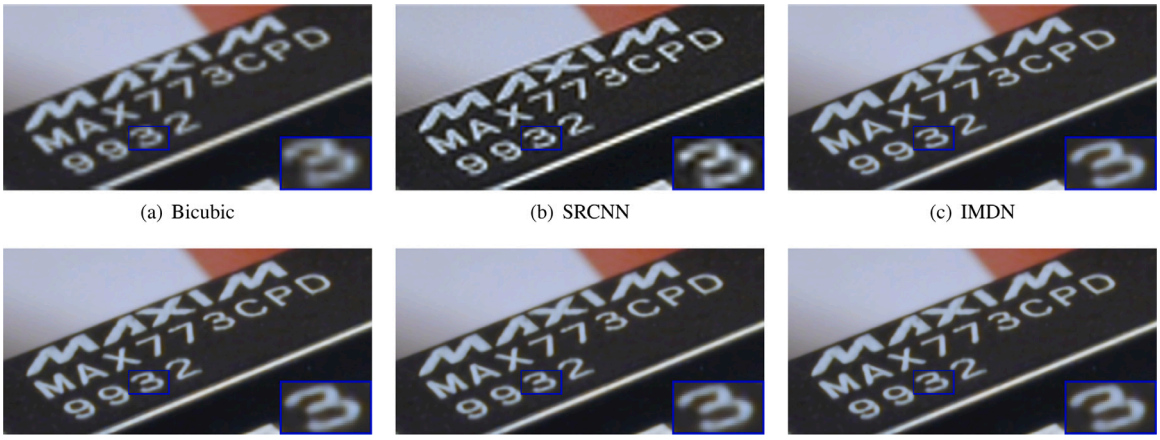


Fig. 4. Visual comparisons for ×4 SR with the BI model on the Set5, B100, Urban100 and Manga109 datasets. The best results are **highlighted**.

Table 4
Computation and parameter comparison (×4 SR on Set5).

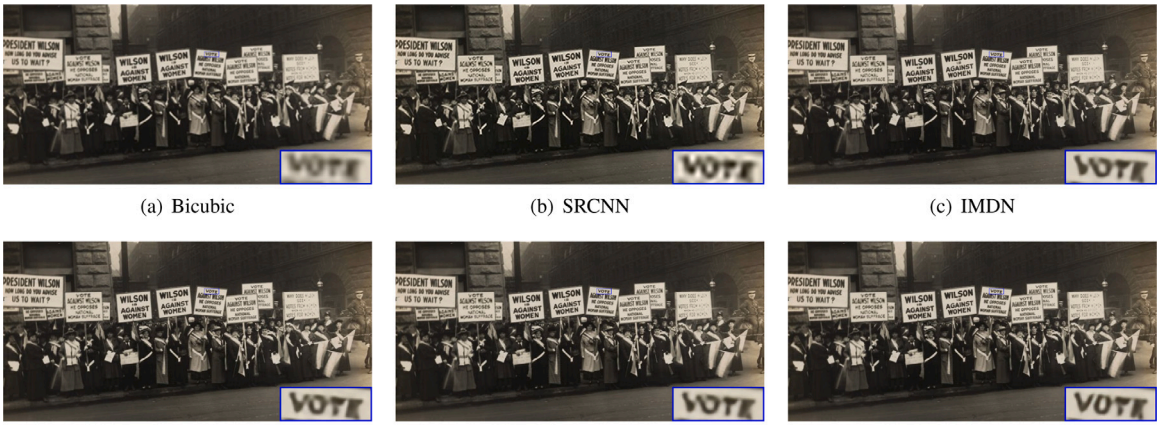
Metric	LapSRN	IMDN	RFDN	LatticeNet	VLESR	LBRN (ours)
Paras (K)	813	715	550	777	331	345
Mult-Adds (G)	149.4	40.9	23.9	43.6	19	20.1
PSNR (dB)	31.54	32.21	32.24	32.18	32.17	**32.33**



(a) Bicubic (b) SRCNN (c) IMDN

(d) RFDN (e) VLESR (f) LBRN (ours)

Fig. 5. Comparison of real-world image "Chip" for ×4 SR.



(a) Bicubic (b) SRCNN (c) IMDN

(d) RFDN (e) VLESR (f) LBRN (ours)

Fig. 6. Comparison of historical image "Historical-006" for ×4 SR.

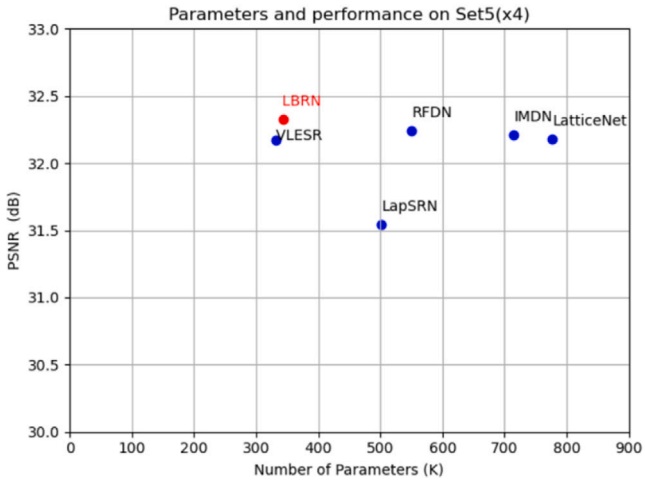


Fig. 7. Performance and the parameters of different methods on Set5 (×4).

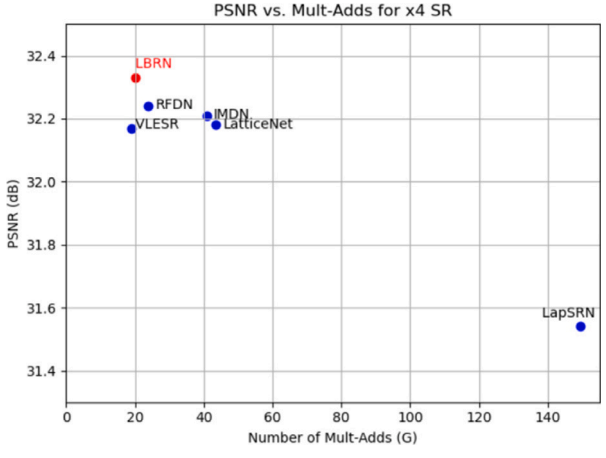


Fig. 8. Performance and the Multi-Adds of different methods on Set5 (×4).

Table 5
Average milliseconds of inference time for various state-of-the-art methods on 100 SR RGB images with 1280 x 720 resolution for ×4 SR.

Method	NGswin	IMDN	RFDN	PAN	VLESR	LBRN (ours)
Time (ms)	156.15	8.56	7.31	12.35	31.19	21.11

we further illustrate the comparison of PSNR vs. Multi-Adds on Set5 for ×4 SR between our LBRN and other lightweight state-of-the-art methods. The comparison results are showed in Table 4, and the corresponding visual illustration is showed in Fig. 8. It can be seen that our LBRN with 20.1 G multi-adds, which is slightly more than 19.0 G multi-adds of VLESR, achieves the highest PSNR 32.33 dB which significantly surpasses the 32.17 dB PSNR of VLESR by 0.16 dB. Compared to other methods including LapSRN, IMDN, RFDN and LatticeNet, our LBRN achieves the highest PSNR with the least cost of Mult-Adds and exceeds the second highest PSNR of RFDN by 0.09 dB. The comparison results demonstrate the effectiveness and efficiency of our proposed LBRN.

4.6. Inference time

In order to evaluate time efficiency even more, we calculate the practical inference times of various state-of-the-art SR methods including NGswin (Choi et al., 2023), IMDN (Hui et al., 2019), RFDN (Liu et al., 2020), PAN (Zhao et al., 2020), and VLESR (Gao & Zhou, 2023). We calculate the average inference times of these comparable methods using their official codes on one hundred super-resolved RGB images with the resolution of 1280 × 720, all in the same experimental setup as our LBRN, and the torch.cuda.Event of PyTorch framework is applied to calculate all inference times. As for ×4 SR, the calculated results of different methods are shown in Table 5. In theory, depth-wise convolution (Haase & Amthor, 2020b, Chollet, 2017) can reduce the number of parameters and computing complexity of standard convolution.

However, the GPU's acceleration performance is currently not able to meet the theoretical value due to the depth-wise convolution's high memory access cost (MAC) to floating-point operations (FLOPs). Therefore, even though our LBRN has fewer parameters and FLOPs theoretically, as Table 5 demonstrates, our method has no obvious advantage over existing lightweight CNN-based models in terms of real inference time. It is noteworthy that our LBRN infers much more quickly than the NGswin (Choi et al., 2023), a transformer-based state-of-the-art method. In addition, our LBRN infers faster than lightweight VLESR (Gao & Zhou, 2023). To further facilitate the inference of lightweight networks, more study is needed to optimize the implementation of depth-wise convolution and speed up its feed-forward process.

5. Conclusions

We propose a lightweight blueprint residual network (LBRN) for lightweight and accurate image SR. Specifically, we propose a novel lightweight blueprint residual block (LBRB) by incorporating the BSConv and the RIR structure. It is extremely lightweight and pays attention to the learning of high-frequency information which helps recover more visual details for lightweight SISR. In addition, we propose an effective feature fusion residual block (EFFRB) which consists of three LBRBs with convolutional kernels of different sizes to learn multi-scale high-frequency information, and the multi-scale high-frequency information is further fused to effectively obtain discriminative feature. Extensive experiments demonstrate the superiority and efficiency of our LBRN on lightweight SISR datasets generated by BI degradation as well as on the real-world photos.

Funding

This work was supported in part by National Natural Science Foundation of China under Grant 62002177, and Natural Science Foundation of Tianjin City, China under Grant 19JCQNJC00300 and 21JCYBJC00110. (Co-author: Jing Xu).

CRediT authorship contribution statement

Fangwei Hao: Methodology, Software, Writing – original draft. **Jiesheng Wu:** Supervision, Visualization. **Weiyun Liang:** Software. **Jing Xu:** Writing – review & editing, Project administration. **Ping Li:** Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Jing Xu reports financial support was provided by National Natural Science Foundation of China. Jing Xu reports financial support was provided by Natural Science Foundation of Tianjin City.

Data availability

The data in the paper are public datasets.

Acknowledgments

The authors would like to thank the reviewers for their insightful comments and useful suggestions.

References

Ahn, N., Kang, B., & Sohn, K.-A. (2018). Fast, accurate, and lightweight super-resolution with cascading residual network. In *Proceedings of the European conference on computer vision* (pp. 252–268).

Arbelaez, P., Maire, M., Fowlkes, C., & Malik, J. (2010). Contour detection and hierarchical image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *33*(5), 898–916.

Bevilacqua, M., Roumy, A., Guillemot, C., & Alberi-Morel, M. L. (2012). *Low-complexity single-image super-resolution based on nonnegative neighbor embedding*. BMVA Press.

Cen, F., Zhao, X., Li, W., & Wang, G. (2021). Deep feature augmentation for occluded image classification. *Pattern Recognition*, *111*, Article 107737.

Choi, H., Lee, J., & Yang, J. (2023). N-gram in swin transformers for efficient lightweight image super-resolution. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 2071–2081).

Chollet, F. (2017). Xception: Deep learning with depthwise separable convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1251–1258).

Chu, X., Zhang, B., Ma, H., Xu, R., & Li, Q. (2021). Fast, accurate and lightweight super-resolution with neural architecture search. In *2020 25th international conference on pattern recognition* (pp. 59–64). IEEE.

Dong, C., Loy, C. C., He, K., & Tang, X. (2015). Image super-resolution using deep convolutional networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *38*(2), 295–307.

Dong, C., Loy, C. C., & Tang, X. (2016). Accelerating the super-resolution convolutional neural network. In *Computer vision–ECCV 2016: 14th European conference, Amsterdam, the Netherlands, October 11-14, 2016, Proceedings, Part II 14* (pp. 391–407). Springer.

Fan, M., Lai, S., Huang, J., Wei, X., Chai, Z., Luo, J., & Wei, X. (2021). Rethinking bisenet for real-time semantic segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 9716–9725).

Gao, Z., Wang, L., Han, B., & Guo, S. (2022). Adamixer: A fast-converging query-based object detector. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 5364–5373).

Gao, D., & Zhou, D. (2023). A very lightweight and efficient image super-resolution network. *Expert Systems with Applications*, *213*, Article 118898.

Guo, Y., Chen, J., Wang, J., Chen, Q., Cao, J., Deng, Z., Xu, Y., & Tan, M. (2020). Closed-loop matters: Dual regression networks for single image super-resolution. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 5407–5416).

Haase, D., & Amthor, M. (2020a). Rethinking depthwise separable convolutions: How intra-kernel correlations lead to improved mobilenets. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 14600–14609).

Haase, D., & Amthor, M. (2020b). Rethinking depthwise separable convolutions: How intra-kernel correlations lead to improved mobilenets. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 14600–14609).

Hao, F., Zhang, T., Zhao, L., & Tang, Y. (2022). Efficient residual attention network for single image super-resolution. *Applied Intelligence*, *52*(1), 652–661.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770–778).

Hendrycks, D., & Gimpel, K. (2016). Gaussian error linear units (GELUS). arXiv preprint arXiv:1606.08415.

Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., & Adam, H. (2017). Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861.

Hoyer, L., Dai, D., & Van Gool, L. (2022). Daformer: Improving network architectures and training strategies for domain-adaptive semantic segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 9924–9935).

Huang, J.-B., Singh, A., & Ahuja, N. (2015). Single image super-resolution from transformed self-exemplars. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 5197–5206).

Hui, Z., Gao, X., Yang, Y., & Wang, X. (2019). Lightweight image super-resolution with information multi-distillation network. In *Proceedings of the 27th acm international conference on multimedia* (pp. 2024–2032).

Hui, Z., Wang, X., & Gao, X. (2018). Fast and accurate single image super-resolution via information distillation network. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 723–731).

Joseph, K., Khan, S., Khan, F. S., & Balasubramanian, V. N. (2021). Towards open world object detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 5830–5840).

Kim, J., Lee, J. K., & Lee, K. M. (2016a). Accurate image super-resolution using very deep convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1646–1654).

Kim, J., Lee, J. K., & Lee, K. M. (2016b). Deeply-recursive convolutional network for image super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1637–1645).

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.

Lai, W.-S., Huang, J.-B., Ahuja, N., & Yang, M.-H. (2017). Deep laplacian pyramid networks for fast and accurate super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 624–632).

Lai, W.-S., Huang, J.-B., Ahuja, N., & Yang, M.-H. (2018). Fast and accurate image super-resolution with deep laplacian pyramid networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *41*(11), 2599–2613.

Lan, R., Sun, L., Liu, Z., Lu, H., Pang, C., & Luo, X. (2020). MADNet: A fast and lightweight network for single-image super resolution. *IEEE Transactions on Cybernetics*, *51*(3), 1443–1453.

Li, Z., Xi, T., Zhang, G., Liu, J., & He, R. (2021). AutoDet: Pyramid network architecture search for object detection. *International Journal of Computer Vision*, *129*, 1087–1105.

Li, X., Zhao, H., Han, L., Tong, Y., Tan, S., & Yang, K. (2020). Gated fully fusion for semantic segmentation. In *Proceedings of the AAAI conference on artificial intelligence*: *vol. 34*, (no. 07), (pp. 11418–11425).

Lim, B., Son, S., Kim, H., Nah, S., & Mu Lee, K. (2017). Enhanced deep residual networks for single image super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops* (pp. 136–144).

Lin, T.-Y., Dollár, P., Girshick, R., He, K., Hariharan, B., & Belongie, S. (2017). Feature pyramid networks for object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2117–2125).

Liu, Z., Mao, H., Wu, C.-Y., Feichtenhofer, C., Darrell, T., & Xie, S. (2022). A convnet for the 2020s. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 11976–11986).

Liu, J., Tang, J., & Wu, G. (2020). Residual feature distillation network for lightweight image super-resolution. In *Computer vision–ECCV 2020 workshops: Glasgow, UK, August 23–28, 2020, proceedings, part III 16* (pp. 41–55). Springer.

Luo, X., Qu, Y., Xie, Y., Zhang, Y., Li, C., & Fu, Y. (2022). Lattice network for lightweight image restoration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *45*(4), 4826–4842.

Ma, W., Xu, S., Ma, W., Zhang, X., & Zha, H. (2022). Progressive feature learning for facade parsing with occlusions. *IEEE Transactions on Image Processing*, *31*, 2081–2093.

Matsui, Y., Ito, K., Aramaki, Y., Fujimoto, A., Ogawa, T., Yamasaki, T., & Aizawa, K. (2017). Sketch-based manga retrieval using manga109 dataset. *Multimedia Tools and Applications*, *76*, 21811–21838.

Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., Lin, Z., Desmaison, A., Antiga, L., & Lerer, A. (2017). Automatic differentiation in Pytorch.

Shi, W., Caballero, J., Huszár, F., Totz, J., Aitken, A. P., Bishop, R., Rueckert, D., & Wang, Z. (2016). Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1874–1883).

Tai, Y., Yang, J., & Liu, X. (2017). Image super-resolution via deep recursive residual network. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 3147–3155).

Timofte, R., Agustsson, E., Van Gool, L., Yang, M.-H., & Zhang, L. (2017). Ntire 2017 challenge on single image super-resolution: Methods and results. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops* (pp. 114–125).

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

Wang, Q., Wu, B., Zhu, P., Li, P., Zuo, W., & Hu, Q. (2020). ECA-Net: Efficient channel attention for deep convolutional neural networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 11534–11542).

Zeyde, R., Elad, M., & Protter, M. (2012). On single image scale-up using sparse-representations. In *Curves and surfaces: 7th international conference, Avignon, France, June 24-30, 2010, Revised Selected Papers 7* (pp. 711–730). Springer.

Zhai, X., Kolesnikov, A., Houlsby, N., & Beyer, L. (2022). Scaling vision transformers. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 12104–12113).

Zhang, Y., Li, K., Li, K., Wang, L., Zhong, B., & Fu, Y. (2018). Image super-resolution using very deep residual channel attention networks. In *Proceedings of the European conference on computer vision* (pp. 286–301).

Zhang, Y., Tian, Y., Kong, Y., Zhong, B., & Fu, Y. (2020). Residual dense network for image restoration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *43*(7), 2480–2495.

Zhang, L., & Wu, X. (2006). An edge-guided image interpolation algorithm via directional filtering and data fusion. *IEEE Transactions on Image Processing*, *15*(8), 2226–2238.

Zhang, K., Zuo, W., & Zhang, L. (2018). Learning a single convolutional super-resolution network for multiple degradations. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 3262–3271).

Zhao, H., Kong, X., He, J., Qiao, Y., & Dong, C. (2020). Efficient image super-resolution using pixel attention. In *Computer vision–ECCV 2020 workshops: Glasgow, UK, August 23–28, 2020, proceedings, part III 16* (pp. 56–72). Springer.

Zheng, A., Zhang, Y., Zhang, X., Qi, X., & Sun, J. (2022). Progressive end-to-end object detection in crowded scenes. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 857–866).