Lightweight Dual-Stream Residual Network for Single Image Super-Resolution

YICHUN JIANG\*, (Graduate Student Member, IEEE), YUNQING LIU, WEIDA ZHAN\*, AND DEPENG ZHU
National Demonstration Center for Experimental Electrical and Electronic Technology, Changchun University of Science and Technology, Changchun, Jilin 130022, China
Corresponding author: Yunqing Liu (liuyunqing@cust.edu.cn)

This work was supported by “13th Five Year Plan” Scientific Research Program of Jilin Provincial Department of Education under Grant JJKH20200783KJ.

ABSTRACT The deep convolutional neural network has achieved great success in the Single Image Super-resolution task. It is obviously that among the well-known super-resolution methods, the deep learning-based algorithms show the most advanced performance. However, the most advanced algorithms currently use complex networks with a large number of parameters, which makes it difficult to apply deep learning algorithms on mobile devices. To solve this problem, we propose a lightweight dual-residual network (LDRN) for single image super-resolution, which has better reconstruction quality than most current advanced lightweight algorithms. Due to its fewer parameters and computational expense, real-time and mobile applications of our networks can be easily realized. On the basis of the residual module, we propose a new residual unit, which uses two depthwise separable (DW) convolution to obtain better balance between feature extraction capacity and lightweight performance. We further design a dual-stream residual block, which contains a multiplication branch and an addition branch. The dual-stream residual block can improve the reconstruction performance more effectively than expanding the network width. In addition, we also designed a new up-sampling module to simplify the previous up-sampling methods. Extensive experimental results show that our network has better reconstruction performance and lightweight performance than most existing state-of-the-art algorithms. Our code is available at https://github.com/Jiangyichun-cust/pytorch-LDRN.

INDEX TERMS Deep convolutional neural networks, single image super-resolution, lightweight model, model compression, residual module, up-sampling module.

I. INTRODUCTION

Single image super-resolution (SISR) is a classic problem in the field of low-level computer vision. Addressing the problem has been proven to be useful in many practical cases, such as medical imaging [1], [2], infrared imaging [3]–[5] and remote sensing imaging [6]–[8]. From mathematical point of view, SISR needs to build a degradation model from high-resolution (HR) images to low-resolution (LR) images and fit the inverse function [9]. The goal of SISR is to generate high-quality super-resolution (SR) images with clear details that are as close as possible to HR images. However, some high-frequency information is lost in the degradation process, so the LR image may correspond to multiple possible HR images, which makes the SISR problem ill-posed.

In order to solve the problem of SISR, many methods have been proposed and proved to be feasible, including interpolation-based [10], reconstruction-based [11], and learning-based algorithms [12]. The interpolation-based algorithms is simple and effective, which are based on the assumption that pixels obey simple functions such as linear, quadratic or cubic functions, only consider the structure information of image in a small range. Therefore, the interpolation-based algorithm cannot reconstruct the high-frequency information of the image, which leads to poor SR image quality. The reconstruction-based algorithms usually require multiple frames of images to provide the necessary information and require artificially well-designed prior knowledge to constrain the reconstruction locally or non-locally. In these methods, when the scale factor becomes larger, the reconstruction performance will drop sharply.

Recently, the mainstream of SISR method is learning-based algorithm, in which the SISR algorithm based on sparse...
representation [12]–[14] shows a better effect than traditional algorithms. It treats LR images as low-dimensional manifolds of HR images, and constructs an overcomplete dictionary through regularization terms. The SISR methods based on Deep Convolutional Neural Networks (DCNN) have shown strong feature representation capabilities in recent years, and have been proven to be able to fit very complex mappings.

As a problem with a high-dimensional solution space, DCNN-based algorithms are very suitable for solving the problem of SISR. At present, DCNN-based super-resolution algorithm provides excellent reconstruction performance in various super-resolution tasks. Dong et al. [15] proposed an SRCNN consisting of only three convolutional layers. Although the number of layers is small, it achieves better reconstructed image quality than most traditional methods. However, only a few convolutional layers that are used in the SISR network using is insufficient for good reconstruction capability. The VDSR proposed by Kim et al. [16], adopted a very deep network structure, proved that the network depth and the number of parameters have obviously better effects on the reconstruction quality in the SISR. Ledig et al. [17] proposed SRRResNet, which has made great progress compared with previous methods, is the first time that ResNet [18] and SISR are combined. Lim et al. [19] studied the influence of some commonly used structures on SISR, and further improved and enhanced the residual networks for SISR called EDSR. Although the DCNN-based algorithm has achieved great success, there are still some problems in most networks:

1) Most networks tend to be designed with more parameters and computational expense to improve the performance of these networks. This means that these methods are difficult to port to mobile devices, and it is now impossible to achieve real-time processing. As a low-level vision task, super-resolution is only a part of image preprocessing, so these inefficient methods are unacceptable in many fields.

2) Most network structures are designed for other tasks, especially image classification and segmentation. If these networks are directly applied to SISR, the expected results may not be achieved. The current residual network structure does not make enough targeted improvements for SISR images, and ignores the complexity of degrading from HR images to LR images.

3) Some widely used structures in SISR can be improved and simplified to reduce unnecessary parameters and computational expense.

In order to solve the above problems, there has been a lot of works in recent years focusing on the realization of lightweight DCNN-based algorithm. A widespread strategy is to share parameters between different blocks, such as Multi-scale Residual Network for Image Super-Resolution (MSRN) and A Fast and Lightweight Network for Single-Image Super Resolution (MADNet) [20], [21]. However, the number of parameters still has a great impact on the performance of the network, which cannot be solved well by sharing parameters. For SISR, the network can better fit the inverse function of the image degradation process and is more conducive to improve the efficiency of the network.

In this article, we propose a lightweight neural network called LDRN with good reconstruction ability to achieve a better balance between performance, parameter volume, and calculation volume. As shown in Figure 1, our network is superior to many state-of-the-art algorithm. In summary, the main contributions in this article are listed as follows:

1) We built an efficient architecture that incorporates multi-level features, adds global skip connections to increase sparsity, and removes all batch normalization (BN) for low-level computer vision like SISR.

2) We propose a new residual unit, which is composed of two depthwise separable convolution (DW convolution) that perform dimension expansion and dimension compression on features respectively. It has good feature extraction capability and fewer parameters.

3) We propose a new dual-stream residual block consisting of dual branches. The additive branch and the multiplicative branch can better fit the inverse process of the image degradation process. As a result, the better reconstruct performance is obtained under the same parameter amount and computational expense.

4) We propose a new up-sampling module, which removes the dimensional expansion before pixel shuffle, and changes to dimensional compression to directly output SR images through pixel shuffle. It significantly reduces the number of parameters with almost no performance loss.

II. RELATED WORK

Deep Convolutional Neural Network has been widely used in many fields [22]–[25] since the success of AlexNet in image classification tasks [26], and Single image super-resolution is one of them. In this paper, we focus on the SISR method based on lightweight neural networks. In addition, due to space limitations, we only discuss two branches: DCNN-based SISR algorithm and model compression.

A. DCNN-BASED SINGLE IMAGE SUPER-RESOLUTION

Recently, deep learning has achieved significant improvements on SISR tasks. SRCNN, proposed by Dong et. al, only uses three convolutional layers to represent Patch extraction and representation, Non-linear mapping and Reconstruction operations. SRCNN interpolates the LR image to the size of the HR image at the beginning, which introduces a lot of redundant information and greatly increases the computational expense. Therefore, FSRCNN was proposed to solve these problems [27]. It processes the features of the same size as the LR image, and then uses Deconvolution to upsample the image at the tail. In subsequent research, the structure of processing under the LR image size and then up-sampling at the tail is widely used, which greatly reduces the computational expense and avoids the misconvergence arising from simply stacking convolutional layers.
Inspired by ResNet, in addition to stacking a large number of convolutional layers, VDSR also adds a global skip connection to solve the vanishing/exploding gradients problem. VDSR also shows that a deeper network can bring better reconstruction results. Kim et al. [28] applied the recurrent neural network to the SISR task and proposed a DRCN with a deep recursive architecture, which can reuse layers to share parameters. SRResnet obtains good feature extraction ability by stacking a large number of residual blocks, and introduces PReLU to learn negative coefficients. EDSR deletes the batch normalization layer in the residual block. This is because experiments have found that the batch normalization layer destroys the absolute difference between pixels, which is unfavorable for low-level computer vision tasks.

On the other hand, LR corresponds to a complex SR solution since SISR is an ill-posed problem. Although these state-of-the-art algorithms have high PSNR/SSIM, traditional DCNN-based algorithm tends to generate the average of these SR solutions [29], so the generated image will be visually blurred to a certain extent. Some researches focus on generating clear and seemingly real images, such as SRGAN, ESRGAN, etc [30]–[32]. They design a discriminator to provide adversarial loss to reduce the gap between the generated data distribution and the real data distribution. Although GAN can generate data that looks realistic enough, the generated data is not true, which is unacceptable in medical, military and other fields. And this will cause the GAN network to perform poorly in PSNR/SSIM.

In this article, we focus on the pixel-level accuracy of the reconstructed image. On the other hand, our goal is to design a network with good reconstruction performance, and achieve real-time and lightweight super-resolution.

B. MODEL COMPRESSION

The lightweight design of neural networks has always been a research hotspot, and we will provide a brief survey of some current model compression methods.

Network pruning reduces the size of the network by deleting some unnecessary branches [33], [34], but it still requires training a large network and designing methods to evaluate the importance of branches. Knowledge distillation requires a large teacher network to guide smaller distilled network training [35], [36]. It also needs to calculate the gap between the two network feature maps, which requires more complex techniques. At present, the most common strategy is to directly design a lightweight deep convolutional neural network. For example, MobileNetV1 proposed DW convolution [37], which uses a convolution kernel to process a feature map, and then uses $1 \times 1$ convolution to linearly fuse the output. Such an operation significantly reduces the amount of parameters and computational expense, with little performance degradation. In order to reduce the amount of model parameters and improve performance, MobileNetV2 introduces a residual structure and bottleneck [38]. A scale factor is set to expand the channel in the compression model. As a result, good performance is obtained while reducing the number of parameters. ShuffleNet replaces the $1 \times 1$
convolution in MobileNet with pointwise group convolutions, and performs channel shuffle after the first group convolution (GConv) in the residual unit to reduce boundary effects [39].

Recently, many new algorithms have appeared for SR-related model compression. Inspired by MobileNet V1, CARN designed a compression strategy and a cascading framework with recursive layers [40]. In addition, CARN-M, as a lightweight version of CARN, achieves higher precision with fewer parameters and less computational expense. The IDN, proposed by Hui et al. [41], combines an enhancement unit and a compression unit in the information distillation module to mix multiple types of features and extract useful information. Since each layer has fewer filters and group convolution is used, it has higher speed and fewer parameters. Li et al. [42] referred to the structure of U-Net, adopted the residual block design of MobileV2, and proposed a Super Lightweight Super-Resolution Network (s-LWSR). s-LWSR also designed an information pool module to avoid overfitting and fully utilize the information.

Although these lightweight networks have achieved impressive results, there is still a lot of room for improvement. We can still build a more efficient lightweight network to better balance the relationship between parameters, computational expense and performance.

### III. PROPOSED METHOD

As we mentioned in Section I, we propose a new lightweight dual-stream residual network for SISR, called LDRN. This section will introduce our proposed LDRN in detail. First, we provide a comprehensive and detailed description of the main network framework of LDRN and its sub-modules, including the dual-stream residual block and up-sampling module we proposed, which are the core of our method. Then, we explained the loss function used in network training.

#### A. NETWORK FRAMEWORK

Before explaining the technical details of these sub-modules, it is necessary to point out that the goal of LDRN is to establish a mapping from low-resolution images $I_{LR}$ to high-resolution images $I_{HR}$, so we formulated the whole process as:

$$I_{SR} = F(I_{LR}).$$

where $I_{SR}$ represents super-resolution images which are the outputs of the network, $F(\cdot)$ represents the mapping relationship from $I_{LR}$ to $I_{HR}$ fitted by networks. As shown in Figure 2, our LDRN mainly consists of four sub-modules: shallow feature extraction module(SFEM), deep feature extraction module(DFEM), feature fuse module(FFM), and up-sampling module(UM).

First, the low-resolution image is processed by SFEM into shallow features as the input for the subsequent DFEM. On the other hand, the outputs of SFEM are saved and then fused with the deep feature in the subsequent process after the DFEM to provide texture information of the image. Such a shallow feature extraction module does not need a complex structure, so we use two convolution layers to extract the shallow feature. If the output channels of DFEM were different from other sub-modules, a $1 \times 1$ convolution block for dimension reduction is added at the tail of SFEM. The operation can be defined as:

$$f_1 = \sigma(w_{3 \times 3} \ast I_{LR}).$$

$$f_2 = \sigma(w_{3 \times 3} \ast f_1).$$

$$f_3 = \begin{cases} \sigma(w_{1 \times 1} \ast f_2), & C_{SFEM} \neq C_{DFEM} \\ f_2 & \text{otherwise} \end{cases}$$

where $f_i$ is the features of $i$-th layer, $\sigma(x) = \max(0, 2x, x)$ represents the LeakyReLU activation function, the $w$ represents the weight of the convolution kernel, $C_m$ is the number of channels of the module $m$. According to our plenty experiments, the Batch Normalization and Bias reduce the performance of networks to varying degrees, so we remove all of them from our networks.

Then, the shallow features are sent into the DFEM which consists of a series of well-designed blocks to extract the deep feature. These blocks are specially designed for super-resolution tasks. They are not only lightweight and fast, but also have good performance. Like the previous works, these blocks are connected in series to increase the capacity and depth of the network to obtain deep image structure information, which can be used to reconstruct high-quality SR images. However, if we only use deep features to reconstruct the SR image, some details in the LR image will be lost. These details will be retained in the shallower features. Therefore, in order to use different levels of features and improve the performance of our network, the dense fusion structure is designed in DFEM. The output of all blocks in DFEM is concatenated and passes through a $1 \times 1$ convolutional layer for information fusion and selection. The operation can be defined as:

$$f_i = F_i(f_{i-1}; \theta), \quad 4 \leq i \leq n.$$  

$$f_{n+1} = \sigma(w_{1 \times 1} \ast F_{cat}(f_4, f_5, \ldots, f_n)).$$

where $F_i(\cdot)$ represents the proposed $i$-th dual-stream residual block, $\theta_4$ is the parameters of the $i$-th dual-stream residual block, $F_{cat}(\cdot)$ represents the operation of concatenate.

Finally, we further use global feature fusion, which means adding the output of SFEM $f_3$ and the output of DFEM $f_{n+1}$. The global skip connection structure helps to ensure that the networks focus on reconstructing the high-frequency information lost in the image degradation process, while ignoring the low-frequency information that is easy to reconstruct and occupies most of the image. After the Global feature fusion, these features are processed by two convolution layers and pixel-shuffle to generate the Super-Resolution images. It is a well-designed up-sampling module, which is different from the traditional up-sampling method. It is faster and
lighter, and suitable for lightweight neural networks. So, Equation 1 can be written as:

\[ I_{SR} = F_{UM}(F_{FFM}(F_{DFEM}(F_{SFEM}(I_{LR}))) + F_{SFEM}(I_{LR}))) \] (7)

where \( F_{UM} \) represents the up-sampling module, \( F_{FFM} \) represents the feature fuse module, \( F_{DFEM} \) represents the deep feature extraction module, and \( F_{SFEM} \) represents the shallow feature extraction module.

**B. DUAL-STREAM RESIDUAL BLOCK**

In our proposed LDRN, the deep features are extracted by a series of fast and lightweight basic residual blocks, which is one of the most important designs. Here, we will provide the detailed description of the structure of our proposed dual-stream residual block.

- **Residual unit:** Our network no longer uses traditional convolution with a large amount of calculations and parameters. The traditional \( 3 \times 3 \) convolution can be replaced by DW Convolution and a closely followed \( 1 \times 1 \) convolution without significant performance degradation. MobileNet V2 has shown through experiments that the width of the DW convolution has a significant influence on the performance of the residual block. Based on extensive experiments, we found that the number of channels is also important for the performance of the network in low-level computer vision tasks. Specifically, expanding dimension in the residual block always performs better than the compressing dimension. Our proposed residual unit used in our dual-stream residual block is shown in Figure 3.

Different from previous works, our LDRN uses two DW convolutions to extract the feature, one of them is used to initially extract features and expand dimension, while the other is used to refine features and compress dimension. \( 1 \times 1 \) convolution performs information exchange between different channels, and its parameter amount is much larger than DW convolution. It is inefficient to use it only as a compress channel. So we insert DW convolution after compressing the dimension to improve the efficiency of the residual unit. Then, we replace all the activation functions in the residual unit with LeakyReLU to learn negative factors. Experiments show that the proposed structure has good feature extraction capability and fewer parameters. For the input \( x_r \), our proposed residual unit \( F_e \) can be defined as:

\[ F_{ru1}(x) = \sigma(dw_{3 \times 3} \ast \sigma(w_{1 \times 1}^{expand} \ast x))) \] (8)

\[ F_{ru2}(x) = \sigma(dw_{3 \times 3} \ast \sigma(w_{1 \times 1}^{compress} \ast x))) \] (9)

\[ F_e = F_{ru2}(F_{ru2}(x)) \] (10)

where \( dw_{3 \times 3} \) represents the weight of DW Convolution, \( w_{1 \times 1}^{expand} \) represents the \( 1 \times 1 \) convolution for expanding dimension, and \( w_{1 \times 1}^{compress} \) represents the \( 1 \times 1 \) convolution for compressing dimension.

- **Dual information stream:** For the SISR task, it has been proved that the residual structure is a good choice because it can ensure that the backward gradient is propagated more effectively. On the other hand, the residual structure allows the network to focus on estimating the high-frequency information of the images, which is lost during image degradation. However, a large number of studies have found that simply stacking residual blocks to build a very deep network would still cause training difficulties. As the network deepens, the performance
increase is also limited. With the massive increase in the number of network parameters and calculations, the performance growth is quite limited. Therefore, it is necessary to design an efficient residual structure.

Base on the proposed residual unit, we propose a dual information stream structure with multiplicative information stream and additive information stream. The traditional residual block and its improved version only learn the residual information (additive difference) of the main information stream. But the degradation process from HR image to LR image is a highly complex process, and we think it is not enough to consider only additive differences. Therefore, the structure we proposed not only learns the residual information but also learns the multiplicative factors. The proposed dual-stream residual block can be defined as:

\[ f_i = f_{i-1} \times F_{e1}(f_{i-1}) + [F_{e2}(f_{i-1}) + f_{i-1}] \]  

(11)

The proposed dual-stream residual block is more efficient than increasing the expansion rate or width under the same amount of parameters, which will be further proved in subsequent experiments.

C. UP-SAMPLING MODULE

Up-sampling LR images to HR images is a crucial step of SISR, which is directly related to the overall structure of the network. The previous research interpolated LR images to HR images at the beginning of the network or upsampled images, so the number of network parameters and calculations increase is also limited. With the massive increase in the number of network parameters and calculations, the performance growth is quite limited. Therefore, it is necessary to design an efficient residual structure. Base on the proposed residual unit, we propose a dual information stream structure with multiplicative information stream and additive information stream. The traditional residual block and its improved version only learn the residual information (additive difference) of the main information stream. But the degradation process from HR image to LR image is a highly complex process, and we think it is not enough to consider only additive differences. Therefore, the structure we proposed not only learns the residual information but also learns the multiplicative factors. The proposed dual-stream residual block can be defined as:

\[ f_i = f_{i-1} \times F_{e1}(f_{i-1}) + [F_{e2}(f_{i-1}) + f_{i-1}] \]  

(11)

The proposed dual-stream residual block is more efficient than increasing the expansion rate or width under the same amount of parameters, which will be further proved in subsequent experiments.

D. LOSS FUNCTIONS

The Loss function of the neural network is used to compare the output of the network with the ground truth. In order to obtain an SR image that is as close as possible to the HR image and has a satisfactory visual effect, we mix the perceptual loss and the pixel loss to become the loss function of LDRN.

First, we used the L1 loss function commonly used in SISR as the pixel loss. The L1 loss function is different from the L2 loss function. It does not over penalize large errors and has different convergence properties. The L1 loss function can be defined as:

\[ \ell_1 = \frac{1}{N} \sum_{p \in I} (x(p) - y(p)). \]  

(15)

where the \( p \) represents the pixel of images, the \( N \) represents the total number of pixels of the HR or SR image. Previous studies have shown that the L1 function is less likely to
fall into local optimal values. However, the L1 function and L2 function suffer from splotchy artifacts in flat areas. Therefore, we add the MS-SSIM loss function as the perceptual loss, which can be defined as:

\[ \ell_{MS-SSIM}(p) = 1 - MS-SSIM(p). \]  

(16)

Our total loss function can be defined as:

\[ \ell_{total} = a \cdot \ell_1 + (1 - a) \cdot \ell_{MS-SSIM}. \]  

(17)

where \( a \) represents the weight used to balance the pixel loss and the perceptual loss. According to the experiment, we set \( a \) to 0.9.

IV. EXPERIMENTS

In section III, we have explained the structure of our proposed network in detail, and we will further prove the performance of the proposed network in this section. We compared the proposed network with the state-of-the-art networks and performed ablation experiments to prove the effectiveness of the proposed component.

A. IMPLEMENTATION DETAILS

- **Dataset:** We train our network on two widely used super-resolution datasets: DIV2K [43] and Flicker2K. DIV2K includes 800 LR-HR image pairs, and Flicker2K includes 2650 image pairs. The HR patch size is cropped to 256 × 256, while the minibatch size is 32. We only flip the image randomly (vertical or horizontal) without any other data enhancement. In order to compare fairly with other methods, we use four widely used data sets: Set5 [44], Set14 [45], BSD100 [46] and Urban100 [47]. Set5 and Set14 are datasets composed of 5 and 14 pictures, and the resolution of these pictures is low. Both BSD100 and Urban100 are datasets composed of 100 high-resolution images, which are more challenging to be super-resolution reconstructed than Set5 and Set14. All images are used to evaluate the performance of the algorithm, and the performance of the algorithm in each dataset is calculated separately.

- **Training Setting:** Since large-scale super-resolution tasks are more challenging, we only focused on 4x up-sampling tasks. So, we used 16 dual-stream residual blocks for the DFEM of LDRN with expanding rate 2. The widths of SFEM, DFEM, FFFM and UM are all set to 64. For implementation, we trained our network with learning rate 0.0001 for all layers for a total of 400 epochs. For optimization, we used Adam with betas to (0.9,0.999) and weight decay 1e-8. In addition, in order to train LDRN better, we have made some fine-tuning and adopted some training techniques such as annealing LR, multi-step LR, etc. We mark the training result as LDRN+. All experiments were conducted using Pytorch on a NVIDIA RTX3090 GPU.

- **Testing Setting:** We mainly focus on the lightweight performance and reconstruction performance of the model. Lightweight performance includes two indicators: parameters and multi-adds, which represent the requirements for memory and computational expense when the network is actually working. In addition, we assume that the output SR image is 1280 × 720 to calculate the multi-adds. The reconstruction performance mainly includes the widely used PSNR and SSIM. We convert the SR image to the YCbCr space according to the convention and then calculate the PSNR and SSIM on the Y channel.

B. COMPARISON WITH STATE-OF-THE-ART MODELS

We compare the performance of our LDRN to 11 state-of-the-art SISR methods: Bicubic interpolation [10], A+ [48], SRCNN, [15], FSRCNN [27], LapSRN [49], CARN-M [40], IDN [41], VDSR [16], MemNet [50], MADNet-LF [21] and s-LWSR32 [42]. We use the source code and weight parameters provided by the authors, or train according to the method provided by the authors to get the result of all methods. Because we concerned more with the challenging 4x super-resolution, we adopted the 4x version of the comparison algorithm. For networks of different sizes like CARN and s-LWSR, we have selected a lightweight version of them for a fair comparison. TABLE 1 shows the SR results on four testing datasets.

It is obvious that our network performed satisfactorily in the four test datasets. The performance of our network on the Set5 and Set14 datasets is not the best. Specifically, the PSNR of our network is lower than several compared algorithms, but SSIM of our network is the best. In view of the small size of both Set5 and Set 14, the results are not accurate enough for measuring network performance persuasively. The B100 and Urban100 datasets have more images and are more difficult to reconstruct. The performance of our network on the B100 and Urban100 datasets is better than all the compared State-of-The-Art algorithms. Our network has significant progress in both PSNR and SSIM compared to previous studies, which demonstrates the superior super-resolution performance of our network. In addition, it is necessary to point out that our network is not the smallest in terms of parameters and Multi-Adds, but it is acceptable on existing mobile devices. Our network can still easily achieve real-time super-resolution without taking up a lot of memory.

In order to better prove the advantages of our proposed method, we train a larger s-LWSR40 on the same datasets and options as our proposed method, which has almost the same parameters and more multi-adds. Under the same conditions, our proposed network has better performance in 4 valid datasets than s-LWSR40.

The results of qualitative analysis strongly prove the advancement and practicality of our network. To further demonstrate the satisfactory super-resolution capability of our algorithm, we conducted a qualitative analysis on the super-resolution reconstruction capability of the network. In addition to Table 1, we show some SR images from the general datasets. For the sake of fairness, the comparison
images we use are from the best super-resolution results published by the original authors. We will not make subjective comparisons for algorithms for which the original author did not give the original image data. These results are raw data without any type of compression. Figures show that our LDRN has a strong reconstruction ability in SISR, whether in terms of the visual authenticity of the image or the accuracy of reconstruction.

Urban100 is one of the most difficult general datasets to reconstruct, so we selected some SR images to show the advanced performance of our proposed algorithm, as shown in Figure 5 and Figure 6. In Figure 5, the guardrail in the SR image of our algorithm does not produce visible blur, which is obviously closer to the HR image than other algorithms. In Figure 6, the building windows reconstructed by our algorithm have a more natural shape than other algorithms. It shows that our network can effectively extract internal information within a specific image.

One of the difficulties of single image super-resolution reconstruction is the Moire pattern in the image degradation process. Whether the details of the image can be reconstructed accurately depends on how much internal information the network extracts from the image and how much external information it learns from the data set. However, due to the lack of depth or parameters of lightweight neural networks, they are often affected by the Moire pattern and cause reconstruction errors. As shown in Fig7, although our network is also a lightweight network like other compared algorithms, because of its good feature extraction ability, it can reconstruct image details more accurately and is less affected by the Moire pattern.

Whether it is qualitative analysis or quantitative analysis, the results show that our method is better than other advanced methods, and the reconstructed SR image is closer to the HR image.

C. ABLATION STUDY

In order to prove the validity of the proposed structure, we conducted ablation experiments. We give the results and analysis of ablation experiments in this section.

First of all, we pay attention to the effectiveness of the Basic Residual Block. We use $E = 1$, $E = 2$, $E = 3$ and $E = 4$ to represent the expansion rate of the feature extraction module in dual-stream residual blocks. At the same time, we replace our proposed residual unit with the residual unit of MobileNetV2 with the expansion rate 2 and mark it as LDRN(M). As shown in TABLE 2, although PSNR and SSIM increase with the increase of the expansion rate, this paper pays more attention to the lightweight and rapid network. The performance gain brought by the excessive expansion rate is not satisfactory. The expansion rate $E = 2$ not only has fewer parameters and MAC, but also can significantly improve the performance, so we adopt the expansion rate $E = 2$. The performance of our residual unit is significantly better than the result of MobileNetV2’s residual unit, which proves that the proposed residual unit can improve the performance of the network.

Then, we replace the multiplicative branch in the dual information flow with an additive branch. As shown in TABLE 3, Although the modified network has the same parameters and MAC, the performance of network on the public datasets is dropped. The visual effect comparison is shown in Figure 8. After replacing the multiplication branch with the addition branch, the reconstruction quality of the network is significantly reduced visually, and there are more reconstruction errors, which is consistent with the quantitative analysis results. It shows that the multiplicative access in the dual information stream proposed by us can help the network to map LR images to HR images more correctly and help improve the quality of SR images.
FIGURE 6. The super-resolution results of "Image002" (Urban100) with scale factor $\times 4$. LDRN recovers vivid texture regions.

FIGURE 7. The super-resolution results of "Image005" (Urban100) with scale factor $\times 4$. LDRN recovers vivid texture regions.

TABLE 2. Result of an ablation study on the effect of residual unit. The evaluation is on the Set5, Set14, BSD100, and urban public datasets.

| Options  | Scale | Params(K) | Multi-Adds(G) | Set5  | Set14  | BSD100 | Urban100 |
|----------|-------|-----------|---------------|-------|--------|--------|----------|
|          |       |           |               | PSNR/SSIM | PSNR/SSIM | PSNR/SSIM | PSNR/SSIM |
| LDRN(M)  | 4x    | 840       | 46.7          | 31.90/0.887 | 28.02/0.776 | 27.50/0.735 | 25.83/0.776 |
| LDRN(E=1)| 4x    | 598       | 32.6          | 31.92/0.889 | 28.00/0.776 | 27.48/0.734 | 25.76/0.773 |
| LDRN(E=2)| 4x    | 858       | 47.8          | 32.00/0.893 | 28.10/0.780 | 27.57/0.738 | 25.92/0.782 |
| LDRN(E=3)| 4x    | 1132      | 63.1          | 32.05/0.893 | 28.13/0.782 | 27.59/0.739 | 25.95/0.784 |
| LDRN(E=4)| 4x    | 1406      | 78.4          | 32.09/0.894 | 28.15/0.783 | 27.61/0.739 | 25.96/0.786 |

Next, we focus on the design of the up-sampling module. We spliced the proposed up-sampling module and various traditional up-sampling modules to the tail of our network and trained them. According to the order of the four modules mentioned in section 3, we denote the four different up-sampling modules as type-a, type-b, type-c and our proposed type-d.
As shown in Table 4, the up-sampling module proposed by us can reduce a large number of parameters and computational expense with only weak performance degradation, which is suitable for a lightweight network.

Finally, we also use our proposed Dual-stream Residual Block and up-sampling Module to replace some of the residual blocks in CARN-M and its original up-sampling Module. As shown in Table 5, the improved CARN-M is marked...
the network to mobile devices.

In this article, we propose a Lightweight Dual Residual Network for Single Image Super-Resolution. Specifically, a new residual unit is designed to efficiently extract image features, which is a simple and effective structure. Based on the residual unit, we put forward the dual-stream residual block. It can better adapt to complex super-resolution tasks to obtain accurate SR images. Additionally, we further optimize the up-sampling module to get a lighter and faster structure which can still obtain high-quality reconstruction results. Extensive experiments show that our model is better than most of the most advanced lightweight SR algorithms, with excellent reconstruction and lightweight performance. In our future work, we will focus on further improving the performance of the network, developing a multi-scale version, and applying the network to mobile devices.

V. CONCLUSION

In this article, we propose a Lightweight Dual Residual Network for Single Image Super-Resolution. Specifically, a new residual unit is designed to efficiently extract image features, which is a simple and effective structure. Based on the residual unit, we put forward the dual-stream residual block. It can better adapt to complex super-resolution tasks to obtain accurate SR images. Additionally, we further optimize the up-sampling module to get a lighter and faster structure which can still obtain high-quality reconstruction results. Extensive experiments show that our model is better than most of the most advanced lightweight SR algorithms, with excellent reconstruction and lightweight performance. In our future work, we will focus on further improving the performance of the network, developing a multi-scale version, and applying the network to mobile devices.

REFERENCES

[1] D. Qiu, L. Zheng, J. Zhu, and D. Huang, “Multiple improved residual networks for medical image super-resolution,” Future Gener. Comput. Syst., vol. 116, pp. 200–208, Mar. 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167739X20303259

[2] H. Temiz and H. S. Bilge, “Super resolution of B-mode ultrasound images with deep learning,” IEEE Access, vol. 8, pp. 78808–78820, 2020.

[3] Y. Zou, L. Zhang, C. Liu, B. Wang, Y. Hu, and Q. Chen, “Super-resolution reconstruction of infrared images based on a convolutional neural network with skip connections,” Opt. Lasers Eng., vol. 146, Nov. 2021, Art.no. 106717. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0143816621001871

[4] J. Du, H. Zhou, K. Qian, W. Tan, Z. Zhang, L. Gu, and Y. Yu, “RGB-IR cross input and sub-pixel upsampling network for infrared image super-resolution,” Sensors, vol. 20, no. 1, p. 281, Jan. 2020. [Online]. Available: https://www.mdpi.com/1424-8220/20/1/281

[5] X. Zhang, C. Li, Q. Meng, S. Liu, Y. Zhang, and J. Wang, “Infrared image super-resolution by combining compressive sensing and deep learning,” Sensors, vol. 18, no. 8, p. 2587, Aug. 2018. [Online]. Available: https://www.mdpi.com/1424-8220/18/8/2587

[6] H. Wu, S. Zhao, J. Zhang, and C. Lu, “Remote sensing image sharpening by integrating multispectral image super-resolution and convolutional sparse representation fusion,” IEEE Access, vol. 7, pp. 46562–46574, 2019.

[7] H. Wang, Q. Hu, C. Wu, J. Chi, and X. Yu, “Non-locally up-down convolutional attention network for remote sensing image super-resolution,” IEEE Access, vol. 8, pp. 166304–166319, 2020.

[8] A. Xiao, Z. Wang, L. Wang, and Y. Ren, “Super-resolution for ‘Jiilin-1’ satellite video imagery via a convolutional network,” Sensors, vol. 18, no. 4, p. 1194, 2018. [Online]. Available: https://www.mdpi.com/1424-8220/18/4/1194

[9] A. Lugmayr, M. Danelljan, L. Van Gool, and R. Timofte, “SRFlow: Learning the super-resolution space with normalizing flow,” in Computer Vision—ECCV 2020, A. Vedaldi, H. Bischof, T. Brox, and J.-M. Frahm, Eds. Cham, Switzerland: Springer, 2020, pp. 715–732.

[10] R. G. Keys, “Cubic convolution interpolation for digital image processing,” IEEE Trans. Acoust., Speech, Signal Process., vol. ASSP-29, no. 6, pp. 1153–1160, Dec. 1981.

[11] J. Sun, J. Zhu, and M. F. Tappen, “Context-constrained hallucination for image super-resolution,” in Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., Jun. 2010, pp. 231–238.

[12] J. Yang, J. Wright, T. Huang, and Y. Ma, “Image super-resolution as sparse representation of raw image patches,” in Proc. IEEE Comput. Vis. Pattern Recognit., Jun. 2008, pp. 1–8.

[13] J. Yang, J. Wright, T. S. Huang, and Y. Ma, “Image super-resolution via sparse representation,” IEEE Trans. Image Process., vol. 19, no. 11, pp. 2861–2873, Nov. 2010.

[14] J. Zhao, H. Hu, and F. Cao, “Image super-resolution via adaptive sparse representation,” Knowl.-Based Syst., vol. 124, pp. 23–33, May 2017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0950705117301132

[15] C. Dong, C. C. Loy, K. He, and X. Tang, “Learning a deep convolutional network for image super-resolution,” in Computer Vision—ECCV 2014, D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, Eds. Cham, Switzerland: Springer, 2014, pp. 184–199.

[16] J. Kim, J. K. Lee, and K. M. Lee, “Accurate image super-resolution using very deep convolutional networks,” in Proc. IEEE Comput. Vis. Pattern Recognit. (CVPR), Jun. 2016, pp. 1646–1654.

[17] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, “Photo-realistic single image super-resolution by a generator adversarial network,” in Proc. IEEE Comput. Vis. Pattern Recognit., Jun. 2018, pp. 176–184.

[18] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2016, pp. 770–778.

[19] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee, “Enhanced deep residual networks for single image super-resolution,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW), Jul. 2017, pp. 136–144.

[20] J. Li, F. Fang, K. Mei, and G. Zhang, “Multi-scale residual network for image super-resolution,” in Proc. Eur. Conf. Comput. Vis. (ECCV), Sep. 2018, pp. 517–532.

TABLE 5. Result of replacing the module in CARN-M with our proposed module. The evaluation is on the Set5, Set14, BSD100, and urban public datasets.

| Options | Scale | Params(K) | Mul-Adds(G) | Set5 | Set14 | BSD100 | Urban100 |
|---------|-------|-----------|-------------|------|-------|--------|---------|
| CARN-M  | 4x    | 412       | 32.5        | 31.92/0.890 | 28.42/0.776 | 27.44/0.730 | 25.62/0.769 |
| CARN-M* | 4x    | 339       | 33.3        | 31.88/0.890 | 28.20/0.779 | 27.51/0.735 | 25.74/0.776 |

FIGURE 9. Comparison of LDRN with or without multiplicative branch.

with ‘*' to distinguished between original CARN-M and improved CARN-M. Although the improved CARN-M significantly reduced the PSNR of Set14, it performed better than the original CARN-M on B100 and Urban100. Moreover, the improved CARN-M has fewer parameters than the original CARN-M. Therefore, this proves that our proposed method can be extended to some existing methods to improve their performance.
[21] R. Lan, L. Sun, Z. Liu, H. Lu, C. Pang, and X. Luo, “MADNet: A fast and lightweight network for single-image super-resolution,” IEEE Trans. Cybern., vol. 51, no. 3, pp. 1443–1453, Mar. 2021.

[22] K. He, G. Gkioxari, P. Dollár, and R. Girshick, “Mask R-CNN,” in Proc. IEEE Int. Conf. Comput. Vis. (ICCV), Oct. 2017, pp. 2961–2969.

[23] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You only look once: Unified, real-time object detection,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2016, pp. 779–788.

[24] H. Li and X.-J. Wu, “DenseFuse: A fusion approach to infrared and visible images,” IEEE Trans. Image Process., vol. 28, no. 5, pp. 2614–2623, May 2018.

[25] T.-T. Nguyen-Quynh, S.-H. Kim, and N.-T. Do, “Image colorization using neighbor embedding,” in Proc. 23rd Brit. Mach. Vis. Conf. (BMVC), 2012, pp. 135.1–135.10.

[26] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet classification with deep convolutional neural networks,” in Proc. Adv. Neural Inf. Process. Syst., vol. 25, 2012, pp. 1097–1105.

[27] C. Dong, C. C. Loy, and X. Tang, “Accelerating the super-resolution convolutional neural network,” in Computer Vision—ECCV 2016, B. Leibe, J. Matas, N. Sebe, and M. Welling, Eds. Cham, Switzerland: Springer, 2016, pp. 391–407.

[28] J. Kim, J. K. Lee, and K. M. Lee, “Deeply-recursive convolutional network,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2016, pp. 1637–1645.

[29] M. Xiao, S. Zheng, C. Liu, Y. Wang, D. He, G. Ke, J. Bian, Z. Lin, and T.-Y. Liu, “Invertible image rescaling,” in Computer Vision—ECCV 2020, A. Vedaldi, H. Bischof, T. Brox, and J.-M. Frahm, Eds. Cham, Switzerland: Springer, 2020, pp. 126–144.

[30] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, “Mobilenetv2: Inverted residuals and linear bottlenecks,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2018, pp. 126–135.

[31] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet classification with deep convolutional neural networks,” in Proc. Adv. Neural Inf. Process. Syst., vol. 25, 2012, pp. 1097–1105.

[32] S. S. Andrei, N. Shapovalova, and W. Mayol-Cuevas, “SUPERVEGAN: A multi-scale adversarial training approach for single-image super-resolution,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jul. 2017, pp. 626–632.

[33] Y. Tai, J. Yang, X. Liu, and C. Xu, “MemNet: A persistent memory network for image restoration,” in Proc. IEEE Int. Conf. Comput. Vis., Oct. 2017, pp. 4539–4547.

[34] M. Adil, S. Mamoon, A. Zakir, M. A. Manzoor, and Z. Lian, “Multi scale-adaptive super-resolution person re-identification using GAN,” IEEE Access, vol. 8, pp. 177351–177362, 2020.

[35] S. A. Aketi, S. Roy, A. Raghunathan, and K. Roy, “Gradual channel pruning while training using feature relevance scores for convolutional neural networks,” IEEE Access, vol. 8, pp. 171924–171932, 2020.

[36] J.-B. Huang, A. Singh, and N. Ahuja, “Single image super-resolution from transformed self-exemplars,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2015, pp. 5197–5206.

[37] B. Li, B. Wang, J. Liu, Z. Qi, and Y. Shi, “S-LWSR: Super lightweight super-resolution network,” in Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit., Jun. 2018, pp. 4548–4556.

[38] G. Hinton, O. Vinyals, and J. Dean, “Distilling the knowledge in a neural network,” 2015, arXiv:1503.02531 [Online]. Available: http://arxiv.org/abs/1503.02531

[39] J.-W. Jung, H.-S. Heo, H.-J. Shim, and H.-J. Yu, “Knowledge distillation in acoustic scene classification,” IEEE Access, vol. 8, pp. 166870–166879, 2020.

[40] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, “MobileNets: Efficient convolutional neural networks for mobile vision applications,” 2017, arXiv:1704.04861 [Online]. Available: http://arxiv.org/abs/1704.04861

[41] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, “MobilenetV2: Inverted residuals and linear bottlenecks,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2018, pp. 4510–4520.

[42] X. Zhang, X. Zhou, M. Lin, and J. Sun, “ShuffleNet: An extremely efficient convolutional neural network for mobile devices,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2018, pp. 6848–6856.

[43] N. Ahn, B. Kang, and K.-A. Sohn, “Fast, accurate, and lightweight super-resolution with cascading residual network,” in Proc. Eur. Conf. Comput. Vis. (ECCV), 2018, pp. 252–268.

[44] Z. Hui, X. Wang, and X. Gao, “Fast and accurate single image super-resolution via information distillation network,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2018, pp. 723–731.

[45] B. Li, B. Wang, J. Liu, Z. Qi, and Y. Shi, “S-LWSR: Super lightweight super-resolution network,” IEEE Trans. Image Process., vol. 29, pp. 8368–8380, 2020.

[46] E. Agustsson and R. Timofte, “NTIRE 2017 challenge on single image super-resolution: Dataset and study,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW), Jul. 2017, pp. 126–135.

[47] R. Zeyde, M. Elad, and M. Protter, “On single image scale-up using sparse-representations,” in Proc. Int. Conf. Curves Surfs. Berlin, Germany: Springer, 2010, pp. 711–730.

[48] D. Martin, C. Fowlkes, D. Tal, and J. Malik, “A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics,” in Proc. 8th IEEE Int. Conf. Comput. Vis. (ICCV), vol. 2, Jul. 2001, pp. 416–423.

[49] J.-B. Huang, A. Singh, and N. Ahuja, “Single image super-resolution from transformed self-exemplars,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2015, pp. 5197–5206.

[50] R. Timofte, V. De Smet, and L. Van Gool, “A+-: Adjusted anchored neighborhood regression for fast super-resolution,” in Proc. Asian Conf. Comput. Vis. Cham, Switzerland: Springer, 2014, pp. 111–126.

[51] W.-S. Lai, J.-B. Huang, N. Ahuja, and M.-H. Yang, “Deep Laplacian pyramid networks for fast and accurate super-resolution,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jul. 2017, pp. 624–632.

[52] Y. Tai, J. Yang, X. Liu, and C. Xu, “MemNet: A persistent memory network for image restoration,” in Proc. IEEE Int. Conf. Comput. Vis., Oct. 2017, pp. 4539–4547.

YICHUN JIANG (Graduate Student Member, IEEE) was born in Shaoguan, Guangdong, China. He received the B.S. degree in electrical engineering and its automation from Changchun University of Science and Technology, where he is currently pursuing the Ph.D. degree. His research interests include deep neural networks and super-resolution.

YUNQING LIU is currently a Professor and a Supervisor of Ph.D. candidates at Changchun University of Science and Technology. His research interests include digital signal processing and computer vision.

WEIDA ZHAN is currently a Professor and a Supervisor of Ph.D. candidates at Changchun University of Science and Technology. His research interests include digital image processing, infrared imaging technology, and automatic target recognition technology.

DEPENG ZHU was born in Wuwei, Gansu, China, in 1996. He received the B.S. degree in electronic and information engineering from Changchun University of Science and Technology, where he is currently pursuing the Ph.D. degree. His research interests include image fusion, image registration, and object detections.