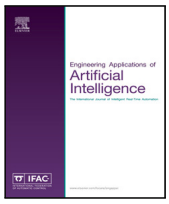
Engineering Applications of Artificial Intelligence 162 (2025) 112595

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

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



Research paper

Large coordinate attention network for lightweight image super-resolution

Fangwei Hao [a,b], Jiesheng Wu [c], Haotian Lu [a,b], Ji Du [a,b], Jing Xu [a,b,*], Xiaoxuan Xu [a,b,*]

[a] College of Artificial Intelligence, Nankai University, Tianjin, 300350, China
[b] Ocean Engineering Research Center, Nankai University, Tianjin 300350, China
[c] the School of Computer and Information, Anhui Normal University, Wuhu 300350, China

ARTICLE INFO

Keywords:
Lightweight image super-resolution
Large coordinate kernel attention module
Large coordinate attention network

ABSTRACT

The Multi-Scale Receptive Field (MSRF) and Large Kernel Attention (LKA) module have been shown to significantly improve performance in the lightweight image super-resolution (SR) task. However, existing lightweight SR methods seldom pay attention to designing lightweight yet effective building block with MSRF for local modeling, and their LKA modules face a quadratic increase in computational and memory footprints as the convolutional kernel size increases. To address the first issue, we propose a simple but effective block, Multi-scale Blueprint Separable Convolutions (MBSConv), as highly efficient building block with MSRF, and it can focus on the learning for the multi-scale information which is a vital component of discriminative representation. As for the second issue, in order to mitigate the complexity of LKA, we propose a Large Coordinate Kernel Attention (LCKA) module which decomposes the two-dimensional convolutional kernels of the depth-wise convolutional layers in LKA into horizontal and vertical one-dimensional kernels. LCKA enables the adjacent direct interaction of local information and long-distance dependencies not only in the horizontal direction but also in the vertical. Besides, LCKA allows for the direct use of extremely large kernels in the depth-wise convolutional layers to capture more contextual information which helps to significantly improve the reconstruction performance, while incurring lower computational complexity and memory footprints. Integrating MBSConv and LCKA, we propose a Large Coordinate Attention Network (LCAN) which is an extremely lightweight SR network with efficient learning capability for local, multi-scale, and contextual information. Extensive experiments show that our LCAN with low model complexity achieves superior performance compared to previous lightweight state-of-the-art SR methods.

1. Introduction

Single image super-resolution (SISR), as a fundamental task in low-level vision, aims to reconstruct a high-resolution (HR) image from its low-resolution (LR) counterpart. This task is inherently challenging due to its ill-posed inverse nature, resulting from massive information loss of degradation. Despite its complexity, SISR holds a wide range of applications across diverse domains, including medical imaging (Feng et al., 2024), remote sensing (Kong et al., 2024), video surveillance (Berardini et al., 2025). In recent years, deep learning-based methods have achieved remarkable progress in this field. Pioneering works such as SRCNN (Dong et al., 2015) and VDSR (Kim et al., 2016a) demonstrate the potential of convolutional neural networks (CNNs) in learning LR-to-HR mappings. Subsequent advancements, including residual learning (He et al., 2016; Lim et al., 2017), dense connection (Zhang et al., 2018b), and attention mechanism (Zhang et al., 2018a), further improve reconstruction accuracy. For instance, EDSR (Lim et al., 2017) and RCAN (Zhang et al., 2018a) leverage deep architectures to achieve state-of-the-art performance. However, these models often suffer from excessive computational complexity, limiting their deployment on resource-constrained devices.

To address efficiency challenges, lightweight SR networks have emerged. Early attempts like DRCN (Kim et al., 2016b) and DRRN (Tai et al., 2017) employ recursive structures to reduce parameters but suffered from performance degradation. Feature distillation strategies, exemplified by IDN (Hui et al., 2018), IMDN (Hui et al., 2019) and, RFDN (Liu et al., 2020b), improve model efficiency through channel splitting and attention mechanisms. In addition, DRSAN (Park et al., 2021) leverages novel dynamic residual attention (DRA) and residual self-attention (RSA) modules for feature enhancement, LatticeNet (Luo et al., 2022) utilizes lattice blocks (LB) and contrastive loss mechanisms, and SCN (Yang et al., 2023) adopts a hierarchical self-calibration module. OSFFNet (Wang and Zhang, 2024) develops an Omni-Stage

* Corresponding authors.
E-mail addresses: haofangwei@mail.nankai.edu.cn (F. Hao), jasonwu@mail.nankai.edu.cn (J. Wu), 2120220499@mail.nankai.edu.cn (H. Lu), 1120230244@mail.nankai.edu.cn (J. Du), xujing@nankai.edu.cn (J. Xu), xuxx@nankai.edu.cn (X. Xu).

https://doi.org/10.1016/j.engappai.2025.112595
Received 9 November 2024; Received in revised form 25 September 2025; Accepted 30 September 2025
Available online 13 October 2025

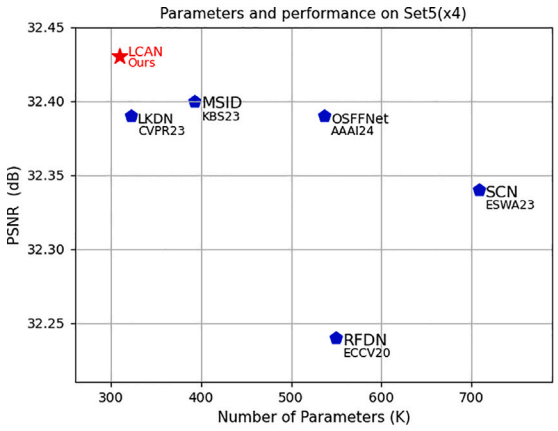


Fig. 1. Comparison of model performance and complexity of different methods on Set5 (Bevilacqua et al., 2012) dataset for 4 × SR. Our model achieves a better performance than previous lightweight state-of-the-art SR methods.

Feature Fusion (OSFF) architecture to integrate features from different levels and capitalize on their mutual complementarity. However, they ignore the effective learning for long-range contextual dependencies, resulting in limited efficiency.

Recently, Large Kernel Attention (LKA) mechanism (Guo et al., 2023) is becoming a common paradigm in CNN-based networks due to its powerful ability for capturing long-range dependencies. Recent lightweight SR methods including MSID (Hu et al., 2023) and LKDN (Xie et al., 2023) apply LKA and its improved version in lightweight SR task, and they use Blueprint Separable Convolutions (BSConv) to alleviate model complexity, achieving effective reconstruction performance. Despite their advancements, two limitations persist: (1) Their BSConv neglects to capture multi-scale information, which is vital for discriminative feature representation; (2) Their LKA modules incur quadratic increases in computational and memory footprints as kernel size increases, hindering their practicality.

This work aims at further bridging the gap between model complexity and reconstruction performance, offering an efficient framework named Large Coordinate Attention Network (LCAN) for lightweight SISR. Specifically, we first design Multi-scale Blueprint Separable Convolutions (MBSConv) as an efficient building block. Unlike standard convolutions, MBSConv integrates multi-scale depth-wise kernels to capture intra-kernel correlations and multi-scale information simultaneously. Secondly, we revisit the limitations of LKA and propose the Large Coordinate Kernel Attention (LCKA) module. By decomposing 2D convolutional kernels into cascaded horizontal and vertical 1D kernels, LCKA enables direct interaction between local details and long-range dependencies in both spatial directions while reducing computational complexity. This design allows the use of extremely large kernels to aggregate contextual information efficiently. Finally, integrating MBSConv and LCKA into a lightweight network, LCAN achieves superior performance with fewer parameters.

When equipped with the proposed MBSConv and LCKA, our LCAN performs favorably against state-of-the-art algorithms for lightweight SR. As shown in Fig. 1, on Set5 (Bevilacqua et al., 2012) dataset, LCAN with the fewest model parameters outperforms recent lightweight SR methods including RFDN (Liu et al., 2020b), DRSAN-32s (Park et al., 2021), MSID (Hu et al., 2023), SCN (Yang et al., 2023), LKDN (Xie et al., 2023), and OSFFNet (Wang and Zhang, 2024).

Overall, our contributions are four-fold:

- We propose a Large Coordinate Attention Network (LCAN) which is an extremely lightweight SR model to recover a high-quality image from the LR input. Besides the effective learning capability for local, multi-scale, and contextual information, our LCAN is more lightweight than previous lightweight SR networks, while achieving superior reconstruction performance.
- We provide Multi-scale Blueprint Separable Convolutions (MBSConv) as a highly efficient building block with multi-scale receptive field for local modeling, and it can focus on the learning for multi-scale information while optimizing intra-kernel correlations compared with standard convolution operation.
- We present a novel Large Coordinate Kernel Attention (LCKA) module to capture more contextual information, while enabling the adjacent direct interaction of local information and long-distance dependencies not only in horizontal direction but also in the vertical. Compared with LKA, the LCKA incurs lower computational complexity and memory footprints, and allows for the direct use of extremely large kernels in the depth-wise convolutional layers, which boost model performance even further.
- We conduct extensive experiments which show that our LCAN with extremely low model complexity achieves superior performance compared with previous state-of-the-art lightweight SR methods.

2. Related work

2.1. Deep networks for lightweight SR

In recent years, image super-resolution based on deep learning has made tremendous progress. The groundbreaking SR network known as SRCNN (Dong et al., 2015), a three-layer convolutional neural network (CNN), can directly simulate the mapping function from LR to corresponding HR. Due to the powerful representation of CNN, SRCNN achieves a significant improvement quantitatively and visually compared to early interpolation-based method (Zhang and Wu, 2006). To further improve the performance, Kim et al. (2016a) design a very deep super-resolution (VDSR) network with 20 convolutional layers. Then, to utilize fewer parameters for large receptive field, Kim et al. (2016b) propose a deep recursive convolutional network (DRCN) by adopting the relatively simple recursive structure. Later, Tai et al. (2017) provide a deep recursive residual network (DRRN), which is an improved one of DRCN. Under the same network depth, DRRN has fewer parameters but outperforms DRCN.

To make a trade-off between model performance and computation, Ahn et al. (2018) propose CARN by combining the group convolutions and the cascading mechanism. Next, Luo et al. (2022) economically adopt two butterfly structures for combining two residual blocks, and they propose a lightweight SR model, LatticeNet, which has relatively low computation and memory requirements. Other lightweight SR networks (Chu et al., 2021; Cheng et al., 2022) are designed automatically by the neural architecture search (NAS) technology which enriches network structures. Another one effective strategy is feature distillation by channel splitting or dimension reduction. Thus, Hui et al. (2018) firstly introduce the feature distillation strategy into the SR task and propose an information distillation network (IDN), which has the advantage of fast execution due to the comparatively few numbers of filters per layer and the use of group convolution. Besides, Hui et al. (2019) improve IDN and propose a lightweight information multi-distillation network (IMDN) by constructing the cascaded information multi-distillation blocks (IMDB), in which the channel splitting strategy is applied multiple times and the channel-wise attention mechanism is introduced. Because of its powerful performance, IMDN wins the first place in the AIM 2019 constrained image SR challenge (Zhang et al., 2019). Moreover, based on IMDN, Liu et al. (2020b) further propose a residual feature distillation network (RFDN) which incorporates their feature distillation connection and shallow residual block. And it achieves good performance while being more lightweight.

Later, BSRN (Li et al., 2022) achieves a new state-of-the-art performance at that time. It utilizes the same feature distillation structure as IMDN and introduces the blueprint separable convolutions

(BSConv) (Haase and Amthor, 2020) to replace the standard convolution in lightweight SR. The results in BSRN show that the BSConv is useful for reducing the parameters of the standard convolution while maintaining effectiveness. However, one BSConv lacks the ability to extract multi-scale information which is a vital component of discriminative representation. To improve the representation capacity, we design multi-scale blueprint separable convolutions (MBSConv) which incorporate the BSConv and the multi-scale structure, and we take it as the highly efficient building block of our LCAN.

2.2. Vision attention

Vision attention can be viewed as an adaptive reweighting according to its input feature, and its superior capability has been demonstrated in not only high-level tasks (e.g., image classification (Hu et al., 2018; Liu et al., 2021a; Sun et al., 2022; Lau et al., 2024), object detection (Liu et al., 2021; Symeonidis et al., 2023; Zhang and Ma, 2023), segmentation (Shi et al., 2023; Lu et al., 2023; Zhang et al., 2023)) but also low-level tasks (e.g., image SR (Park et al., 2021; Luo et al., 2022; Yang et al., 2023; Hao et al., 2024; Hu et al., 2023; Xie et al., 2023)). Since the channel attention mechanism proposed in Hu et al. (2018) shows its effectiveness on image classification, it and its modified ones (Wang et al., 2020; Zhang and Yang, 2021) are quickly introduced and applied to SR networks (Zhang et al., 2018a; Gao et al., 2022; Hao et al., 2022). Although the channel attention mechanism only refines the feature maps along the channel dimension, it shows significant improvements in reconstruction performance. To further improve the discriminative ability for different spatial locations, (Hu et al., 2019; Liu et al., 2020a) incorporate channel-wise attention with spatial attention to force model to automatically learn the multi-level feature maps in global and local manners. Liu et al. (2020c), Zhao et al. (2020) respectively propose one unified attention module by integrating the channel-wise and spatial attention, and the generated three-dimension attention matrix is used to recalibrate feature maps in pixel-level. Besides, (Park et al., 2021) also adopts such joint attention, including channel-wise attention and spatial attention, to exploit discriminative representation of residual features.

After the visual attention network (VAN) (Guo et al., 2023) with large kernel attention (LKA) module is proposed and shows its effectiveness on various tasks, Xie et al. (2023) combine the large kernel attention (LKA) and the Adan optimizer (Xie et al., 2024) to further propose a large kernel distillation network (LKDN), which pushes the model performance of lightweight SR to a new state-of-the-art. Although LKA can provide remarkable performance improvement, it encounters a quadratic increase in computational and memory footprints as the convolution kernel size increases. Recently, Lau et al. (2024) propose a family of large separable kernel attention (LSKA) modules for object recognition, object detection, semantic segmentation, and robustness tests, yet no work is proposed to investigate the effect of LSKA on low-level visual tasks (e.g., image SR). By decomposing the two-dimensional convolutional kernel of the depth-wise convolutional layers into cascaded horizontal and vertical one-dimensional kernels, the LSKA incurs lower computational complexity and memory footprints than LKA, while providing similar performance. However, such decomposition ignores adjacent direct interaction of local information and long-distance dependencies in respective directions, leading to limited performance. Inspired by these works (Guo et al., 2023; Xie et al., 2024; Lau et al., 2024), we propose a large coordinate kernel attention (LCKA) module with adjacent direct interaction of local information and long-distance dependencies in horizontal and vertical directions for the SR task.

3. Method

In this section, we firstly introduce the overall network architecture of LCAN. Then, we give a detailed introduction to the designed multi-scale blueprint separable convolutions (MBSConv), followed by the details of the proposed novel large coordinate kernel attention (LCKA) module. Next, we introduce the proposed multi-scale attention residual block (MARB) in detail. Finally, we show the details of the loss function.

3.1. Network architecture

As shown in Fig. 2, similar to lightweight SR networks (Xie et al., 2023; Li et al., 2022), our LCAN mainly consists of four parts: shallow convolution operation for shallow feature extraction, multi-scale attention residual blocks (MARBs) to extract deep feature, the feature fusion part, and the reconstruction block. Given the input LR image I_{LR} of our LCAN, we firstly replicate the input image I_{LR} n times and concatenate the replicated images along the channel dimension to get I^n_{LR} as the input of the network. Then, we use only one MBSConv, which is composed of a 1×1 convolutional layer and a multi-scale depth-wise convolution layer, to extract the multi-scale shallow feature

$$F_0 = H_s(I^n_{LR}), \tag{1}$$

where $H_s(\cdot)$ is the multi-scale blueprint separation convolutions (MBSConv) operation. Subsequently, M stacked MARBs gradually refine the F_0 by processing each of its input features and producing the refined one. The final MARB's output feature is the obtained deep feature in the part of deep feature extraction. The process of deep feature extraction can be formulated as

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

where $H_k(\cdot)$, F_{k-1}, and F_k denotes the kth MARB operation, its input feature, and the output refined feature, respectively. After gradually refined by M MARBs, M refined features $(F_1, \ldots, F_M)$ from M MARBs are concatenated along the channel dimension, and one 1×1 convolution layer is used to fuse the concatenated feature maps, followed by a MBSConv for smoothing. The process can be formally expressed as

$$F_{fused} = H_{fusion}(\mathrm{Concat}(F_1, \ldots, F_M)), \tag{3}$$

where $\mathrm{Concat}(\cdot)$ denotes the concatenation operation along the channel dimension, $H_{fusion}(\cdot)$ represents the feature fusion function which is made up of a 1×1 convolution layer, a GELU (Hendrycks and Gimpel, 2016) activation, and one MBSConv block, and F_{fused} is the obtained fused feature. At last, a long skip connection is involved across M MARBs, and the result $F_{fused} + F_0$ will be further processed by the reconstruction block, which consists of a 3×3 convolution layer and a sub-pixel convolution layer (Shi et al., 2016) at the tail. We can formally express the reconstruction block as

$$I_{SR} = R(F_{fused} + F_0), \tag{4}$$

where $R(\cdot)$ represents the reconstruction function of the reconstruction block, and I_{SR} is the reconstruction output of the entire network. The detailed configurations of the proposed LCAN are shown in Table 1.

In order to fairly compare with other state-of-the-art SR methods including LKDN, MSID, DRSAN-32s and LatticeNet-CL, we also adopt the L1 loss function to optimize the SR model. Hence, our LCAN's loss function is formulated as

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

Where $H_{\mathrm{LCAN}}(\cdot)$ is the function of our proposed LCAN, and $\ominus$ represents its learnable parameters. In order to make a fair and comprehensive comparison with LKDN, the network is also optimized by Adan (Xie et al., 2024) optimization algorithm in which the adaptive optimization, decoupling weight attenuation, and modified Nesterov impulse are combined.

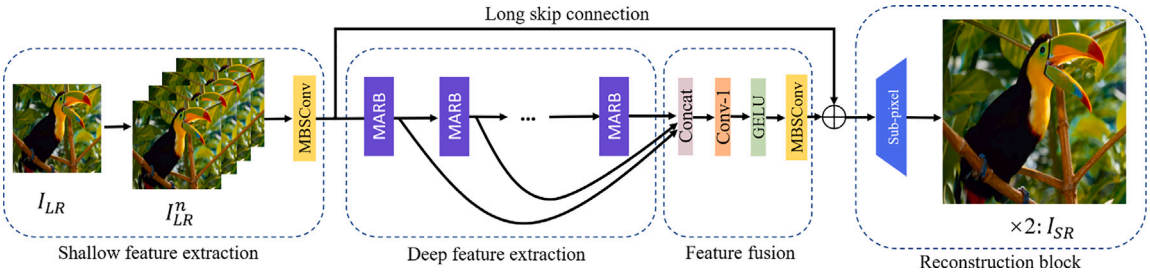


Fig. 2. Network architecture of our LCAN for ×2 SR.

Table 1
Configurations of the proposed LCAN.

Stage	Layer specification	Kernel size	Number
Shallow feature extraction	MBSConv	1 × 1, 3 × 3, 5 × 5	1
Deep feature extraction	MARB	1 × 1, 3 × 3, 5 × 5	8
Feature fusion	Concat	–	1
	1 × 1 convolution	–	1
	GELU	–	1
	MBSConv	1 × 1, 3 × 3, 5 × 5	1
Reconstruction block	Sub-pixel convolution	1 × 1, 3 × 3	1

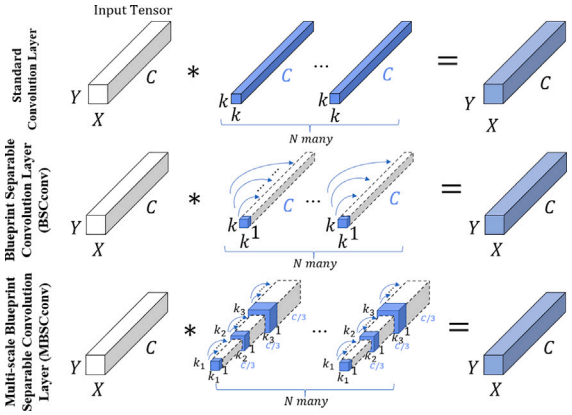


Fig. 3. The comparison of standard convolution layer, BSConv and the proposed MBSConv.

3.2. Multi-scale blueprint separable convolutions (MBSConv)

Recently, a blueprint separation convolution (BSConv) (Haase and Amthor, 2020) is improved from the depth-wise separable convolution (DSConv) (Howard et al., 2017), showing its higher learning efficiency in SR networks (Hu et al., 2023; Xie et al., 2023; Li et al., 2022) for optimizing intra-kernel correlations when compared with standard convolution operation. However, one BSConv lacks the capability of extracting multi-scale information which is crucial to discriminative feature. As demonstrated in Hu et al. (2023), benefiting from the multi-scale receptive field, the acquired multi-scale information is an essential component for effective lightweight SR reconstruction. Nevertheless, existing lightweight SR methods seldom pay attention to designing efficient building block with multi-scale receptive field for local modeling. Inspired by these principles, we design the multi-scale blueprint separable convolutions (MBSConv), which incorporates the BSConv with the multi-scale structure, to extract discriminative feature with multi-scale information, and we take the MBSConv as the efficient building block of our LCAN. MBSConv can be a superior alternative to BSConv in domains where more multi-scale information is needed.

As Fig. 3 shows, one MBSConv is composed of a 1 × 1 standard convolutional layer and a multi-scale depth-wise convolution layer with different kernel sizes. Compared to original BSConv in Haase and Amthor (2020), our MBSConv can not only effectively formulate intra-kernel correlations and allow for a more efficient separation of regular convolutions, but also pay attention to the learning of multi-scale information for more discriminative feature. Considering the complexity in depth-wise convolution layer, we utilize the convolutional kernels with 1 × 1, 3 × 3, and 5 × 5. During the operation of one MBSConv, the input feature F_{in} is firstly refined by a 1 × 1 standard convolutional layer, then the refined feature maps F_{refined} are evenly divided into 3 parts along the channel dimension, and the divided ones are respectively processed by 1 × 1, 3 × 3, and 5 × 5 depth-wise convolutional kernels to generate the corresponding feature maps $F_{1\times1}$, $F_{3\times3}$, and $F_{5\times5}$. At last, $F_{1\times1}$, $F_{3\times3}$, and $F_{5\times5}$ are concatenated along the channel dimension to get feature F_{MBSConv} which is the output of the entire MBSConv.

3.3. Large coordinate kernel attention (LCKA) module

Before introducing the LCKA in detail, we simply revisit the large kernel attention (LKA) module in Guo et al. (2023). Specifically, a 13 × 13 convolution can be decomposed into a 5 × 5 depth-wise convolution (DW-Conv5), a 5 × 5 depth-wise dilation convolution with dilation rate 3 (DW-D-Conv5), and a point-wise convolution (Conv1). And it can be expressed as

$$H_{LKA13}(F^i_{in}) = H_{C1}\left(H_{DDC5}\left(H_{DC5}(F^i_{in})\right)\right) \bigotimes F^i_{\text{in}}, \tag{6}$$

Where $H_{LKA13}(\cdot)$ is the function of LKA, $H_{DDC5}(\cdot)$, $H_{DC5}(\cdot)$, and $H_{C1}(\cdot)$ are the convolutional operations of DW-D-Conv5, DW-Conv5, and Conv1, respectively. Besides, F^i_{in} denotes the input feature of ith LKA module, and ⊗ denotes element-wise product. In addition, we further revisit the key properties of LKA. As shown in Fig. 4-(a), within one LKA module, the spatial local convolution (depth-wise convolution) is mainly used to extract local information, then the extracted local information is immediately processed by the spatial long-range convolution (depth-wise dilation convolution) to capture long-range dependencies. Therefore, the adjacent direct interaction of local information and long-distance dependencies is a key property of LKA. In one LKA module, we can see that the complexity encounters a quadratic increase in computational and memory footprints as the kernel sizes of the depth-wise convolutions increase.

Therefore, how to keep the advantages of LKA including large receptive field and high effectiveness while incurring lower computational complexity and memory footprints, is a key principle for building a

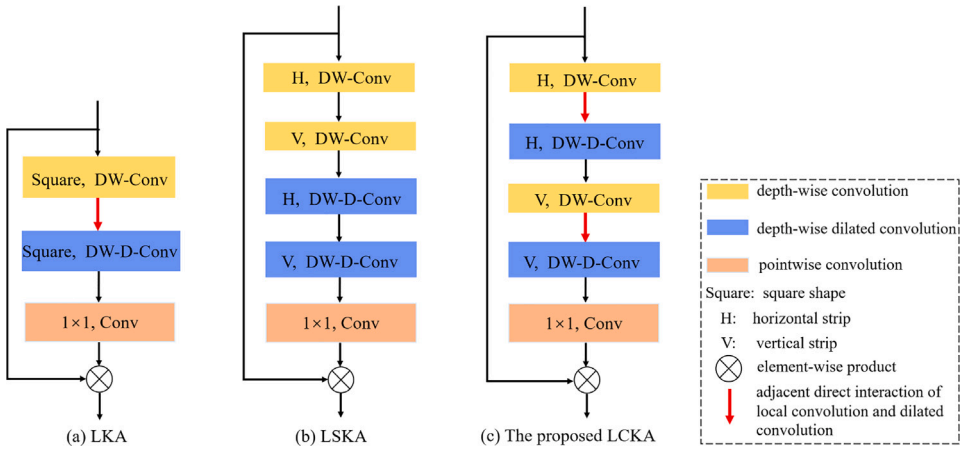


Fig. 4. Comparison on different designs of large kernel attention module. (a) LKA for SR tasks. (b) LSKA in Lau et al. (2024) for high-level tasks (e.g., image classification). (c) The proposed LCKA for low-level tasks (e.g., SR).

lightweight SR module. As Fig. 4-(b) shows, one method in Lau et al. (2024) is to decompose each depth-wise two-dimensional convolutional kernel of $H_{DW-D-Conv5}(\cdot)$ and $H_{DW-Conv5}(\cdot)$ into staggered-connected horizontal and vertical one-dimensional kernels, and the decomposed attention module is denoted as LSKA. Although the LSKA incurs lower computational complexity and memory footprints than LKA, its decomposition ignores the adjacent direct interaction of local information and long-distance dependencies in respective directions, leading to limited performance.

To tackle the issue, we decompose both $H_{DW-D-Conv5}(\cdot)$ and $H_{DW-Conv5}(\cdot)$ into adjacent horizontal one-dimensional convolutional layers and the adjacent vertical ones, that is, the convolutional operations in each direction consist of adjacent layers of a depth-wise one-dimensional convolution layer and a depth-wise dilated one-dimensional convolution layer. Considering the obtained horizontal large kernel and vertical large kernel by decomposition, we denote the entire proposed module as large coordinate kernel attention (LCKA) module, and use it as the building attention module of our LCAN.

In addition, as Fig. 4 shows, due to decomposing large convolution kernels in LKA into horizontal and vertical one-dimensional kernels, both LSKA and LCKA provide significant reductions in computational complexity and memory footprints, but LCKA enables the adjacent direct interaction of local information and long-distance dependencies not only in the horizontal direction but also in the vertical. We will experimentally show that the LCKA is more effective than LSKA for the SR task.

3.4. Multi-scale attention residual block (MARB)

Feature distillation strategy by channel splitting or dimension reduction in Hui et al. (2018, 2019), Liu et al. (2020b), has been demonstrated effective for the lightweight SR task. Following these works, we also use the distillation structure as the basic component of one block. Besides, in order that one block can extract more discriminative representation, we further combine it with the MBSConv and LCKA to propose a powerful multi-scale attention residual block named MARB.

In the kth MARB, the input feature F^{k-1}_{MARB} is refined gradually. As shown in Fig. 5, the process of distillation module based on MBSConv can be formulated as

$$\begin{aligned} F_{d_1}, F_{s_1} &= D_1(F^{k-1}_{\text{MARB}}), S_1(F^{k-1}_{\text{MARB}}), \\ F_{d_2}, F_{s_2} &= D_2(F_{r_1}), S_2(F_{r_1}), \\ F_{d_3}, F_{s_3} &= D_3(F_{r_2}), S_3(F_{r_2}), \\ F_{d_4} &= D_4(F_{r_3}), \end{aligned} \tag{7}$$

where $D_i(\cdot)$ denotes the ith 1×1 convolution of distillation operation, and $S_i(\cdot)$ is the corresponding operation of MBSConv for refinement. F_{d_i} and F_{s_i} respectively represents the obtained ith distilled features and corresponding refined features. During the feature fusion stage, all of the distilled features generated by previous distillation operations are concatenated together, and the concatenated features are fused using a 1×1 convolution. We can express the process as

$$F_{fusion} = H_{fusion}(\text{Concat}(F_{d_1}, F_{d_2}, F_{d_3}, F_{d_4})), \tag{8}$$

where $H_{fusion}(\cdot)$ is the operation of the 1×1 convolution layer, and F_{fusion} denotes the obtained fused feature. Then, to further enhance the learning for discriminative representation, we utilize the proposed LCKA module to capture long-range dependencies. The process of LCKA can be formulated as

$$F_{LCKA} = H_{LCKA}(F_{fusion}), \tag{9}$$

where $H_{LCKA}(\cdot)$, F_{LCKA} represent the function of LCKA, its refined features, respectively. Subsequently, F_{LCKA} is further refined by a 1×1 convolution layer, followed by a pixel normalization (Zhou et al., 2022) module to ease stable model training. We can formulate the process as

$$F_{\text{normalized}} = \text{Norm}_{\text{pixel}}(Conv_{1\times1}(F_{LCKA})), \tag{10}$$

where $Conv_{1\times1}(\cdot)$, $\text{Norm}_{\text{pixel}}(\cdot)$, and $F_{\text{normalized}}$ denote the function of the 1×1 convolution layer, the operation of pixel normalization, and the normalized features, respectively. Finally, we adopt a long skip connection to force the module to focus on learning the residual information. The process can be formulated as

$$F^{k}_{\text{MARB}} = F_{\text{normalized}} + F^{k-1}_{\text{MARB}}, \tag{11}$$

where F^{k}_{MARB} represents the output of the kth MARB.

As we can see in Fig. 5, a MARB has the advantages of feature distillation strategy, capturing multi-scale information, and learning long-range dependencies. Its high efficiency will be demonstrated in the subsequent experimental section.

4. Experiments

In this section, the experimental settings is firstly introduced in detail, and then a series of ablation experiments on LCAN are conducted to verify the efficiency. Next, we compare our LCAN with many other state-of-the-art lightweight SR methods quantitatively and visually. Finally, a complexity analysis of LCAN is performed.

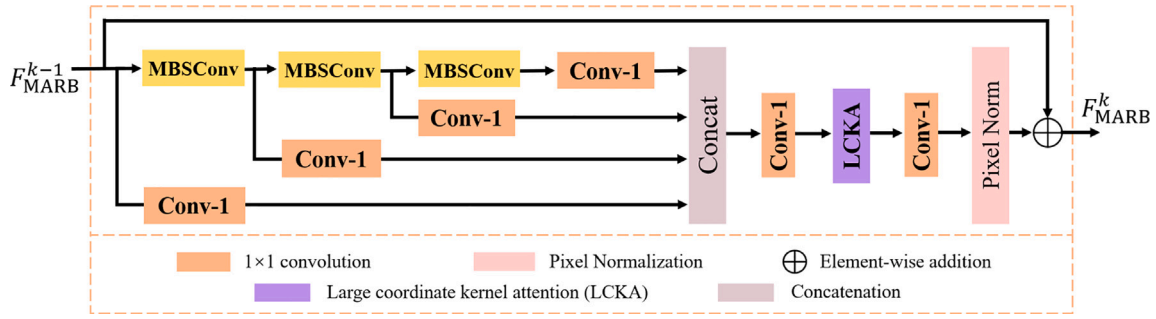


Fig. 5. The details of the proposed MARB.

4.1. Settings

Following the previous advanced works (Hu et al., 2023; Xie et al., 2023; Li et al., 2022), we use the common DF2K (800 images from DIV2K (Timofte et al., 2017) and 2650 images from Flickr2K (Lim et al., 2017)) as training dataset. Following previous methods (Li et al., 2022; Xie et al., 2023; Wang and Zhang, 2024), the proposed LCAN also consists of 8 main modules, i.e., MARBs, and its distillation structure channel number and attention module channel number are set to 55. After finishing training, five widely used benchmark datasets: Set5 (Bevilacqua et al., 2012), Set14 (Zeyde et al., 2012), BSD100 (Arbelaez et al., 2010), Urban100 (Huang et al., 2015) and Manga109 (Matsui et al., 2017) are chosen as test datasets, and the output results of the SR models are converted to YCbCr space. Then, the evaluation metrics: peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM) (Wang et al., 2004) on the Y channel are adopted, and the evaluated quantitative results are applied to showing the model performance. In order to conduct a fair comparison with earlier SR methods, we train our model using the BI training dataset, which is produced via bicubic (BI) degradation using scaling factors of ×2, ×3, and ×4.

During training process, following (Xie et al., 2023), the ADAN (Xie et al., 2024) optimizer with $\beta_1 = 0.98$, $\beta_2 = 0.92$ and $\beta_{13} = 0.99$ is utilized to optimize our SR model from scratch, and the exponential moving average (EMA) is set to 0.999 to stabilize training. For each LR input, the batch size and input patch size are set to 64 and 48 × 48, respectively. We conduct all experiments using the Pytorch (Paszke et al., 2017) framework with an NVIDIA 3090 GPU. The network is trained for 1×10^6 iterations, and the learning rate is set to a constant 5×10^{-3}.

For testing, We conduct the test experiments on both BI test datasets (Set5, Set14, BSD100, Urban100 and Manga109) and Real-World Photos with compression artifacts (Lai et al., 2018; Zhang et al., 2018c).

4.2. Ablation studies

We carry out a series of experiments to analyze the performance of the proposed LCAN and the efficiency of its components including MBSConv and LCKA. Besides, we also conduct experiments to compare the performance of LCKA with the results of LSKA.

Break-down ablation. Firstly, we train the model without the MBSConv, LCKA, and LSKA, on the DF2K dataset for ×4 SR, and take the tested performance 32.23 dB PSNR on the Set5 dataset as the baseline.

Then, we conduct experiments on the same datasets to train and test the model with the MBSConv only. Compared with baseline model, it achieves 0.04 dB gain reaching 32.27 dB PSNR, demonstrating that our MBSConv is effective for extracting feature with multi-scale information and can effectively improve the model performance.

Next, the model with MBSConv and LCKA is trained to verify overall performance of LCAN, and the obtained test result reaches 32.43 dB, which shows the remarkable performance of proposed LCAN for lightweight SR.

Table 2
Investigations of MBSConv, LCKA, and LSKA on the Set5 (×4). We **highlight** the best result.

Baseline	✓	✓	✓	✓
MBSConv	✗	✓	✓	✓
LCKA	✗	✗	✗	✓
LSKA	✗	✗	✓	✗
PSNR/dB	32.23	32.27	32.40	**32.43**

Finally, for comparison, we carry out the same experiments on the model with MBSConv and LSKA, and corresponding test result is 32.40 dB PSNR, which is 0.03 dB lower than the result of the model with MBSConv and LCKA. According to the testing data, the best PSNR of 32.43 dB is achieved by the model with MBSConv and LCKA, proving the effectiveness and superiority of the suggested MBSConv and LCKA. Specifically, our MBSConv can enhance the model performance by capturing discriminative feature with multi-scale information, and our LCKA can capture more contextual information by achieving extremely large receptive field with the direct interaction between local information and long-distance dependencies, which further improves model performance. All experimental results are presented in Table 2.

Ablation of MBSConv. We conduct experiments to validate the selection of the kernel sizes of multi-scale depth-wise convolutions in MBSConv. The input feature maps of MBSConv have 55 channels and they are almost evenly divided into several parts along the channel dimension for corresponding multi-scale depth-wise convolutions. Considering the model complexity, we adopt different multi-scale kernel sizes of the depth-wise convolution in MBSConv, i.e., a set of 1 × 1 (for 18 channels), 3 × 3 (for 18 channels) and 5 × 5 (for 19 channels), a set of 1 × 1 (for 13 channels), 3 × 3 (for 13 channels), 5 × 5 (for 13 channels) and 7 × 7 (for 16 channels), and a set of 3 × 3 (for 18 channels), 5 × 5 (for 18 channels) and 7 × 7 (for 19 channels). Moreover, we also utilize different single-scale BSConvs for comparison experiments. We separately train the models under the same experimental setting and test them on the Urban100 (Huang et al., 2015) dataset for ×4 SR. The detailed results are shown in Table 4.

In one MBSConv, the depth-wise 1 × 1 convolution kernels are less effective than larger ones, and using MBSConv with 3 × 3, 5 × 5 and 7 × 7 depth-wise convolutional kernels can lead to better performance yet result in more model parameters.

In addition, as the depth-wise convolutional kernel size of BSConv increases, it improves model reconstruction performance accordingly, yet the improvement gradually becomes saturated.

Overall, under almost the same number of model parameters, the model with MBSConv can achieve superior performance than the model with BSConv, thanks to the capture of multi-scale information. Due to our aim of using a more lightweight module as a highly efficient building block for lightweight SR, we adopt the MBSConv with 1 × 1,

Table 3
The model performance with different channel numbers on Set5 dataset for ×4 SR. The performance of our model with 55 channels is underlined.

Channel number	32	48	55	64
Model Params (K)	125	247	309	413
PSNR on Set5 (×4)	32.13	32.31	<u>32.43</u>	32.45

3 × 3 and 5 × 5 depth-wise convolutional kernels in our LCAN for a better trade-off of module complexity and performance.

Ablation of LCKA. In order to further verify the effectiveness of the proposed LCKA, we conduct experiments to analyze the effect of different strip sizes of depth-wise convolutions with corresponding dilation rate (dr) in LCKA. Specifically, as strip size k changes from 5 to 9, the dilated rate (dr) which is equal to $(k+1)/2$ varies from 3 to 5. After finishing training separately under the same experimental setting for ×4 SR, the models are tested on the Manga109 (Matsui et al., 2017) dataset. Table 5 shows the detailed experimental results. In addition, we also analyze the effect of different decoupling path settings. Concretely, DWC-DWDC-PC denotes the decoupling order of a depth-wise convolution (DWC), a depth-wise dilation convolution (DWDC), and a point-wise convolution (PC); DWDC-DWC-PC denotes the decoupling path of a depth-wise dilation convolution (DWDC), a depth-wise convolution (DWC), and a point-wise convolution (PC). Notably, for the input feature maps, the module with DWDC-DWC-PC discards the pixels in dilated regions in the process of DWDC, leading to the loss of some contextual information of the input feature maps.

By contrast, when the decoupling order is DWC-DWDC-PC, we can see that as the strip size k increases, the PSNR value of model performance increases from 30.76 dB to 31.03 dB. For the input feature maps, although the LCKA with DWC-DWDC-PC discards the pixels in dilated regions in the process of DWDC, the input feature maps have already all been learned by the DWC, resulting in capturing more complete contextual information via LCKA. Therefore, the PSNR values of the model with the decoupling path DWDC-DWC-PC are lower than aforementioned performance of DWC-DWDC-PC. This means that the decoupling order DWC-DWDC-PC in LCKA is a more effective choice for lightweight SR. Considering the module complexity, we employ the LCKA with the strip size k as 9 in our LCAN to capture the vital long-range contextual information.

In addition, we further analyze and validate the modules with different decoupling path of horizontal and vertical depth-wise convolutions. The H-DWC, H-DWDC, V-DWC, V-DWDC, and PC respectively denote the horizontal depth-wise convolution, horizontal depth-wise dilation convolution, vertical depth-wise convolution, vertical depth-wise dilation convolution, and point-wise convolution. Table 6 shows the detailed validation results, which demonstrate that the LCKA with the decoupling path of H-DWC, H-DWDC, V-DWC, V-DWDC, PC is a superior choice for SR.

The effect of the channel number. In order to validate the effect of the number of feature channels on the model performance, we conduct different validation experiments for ×4 SR. Specifically, we train and test the models with different channel numbers of 32, 48, 55, and 64, respectively. The detailed results are shown in Table 3. It can be seen that as the number of feature channels increases, the performance PSNR of the model increases accordingly. When the number of feature channels increases from 55 to 64, the model performance increases slightly, indicating that the model performance is becoming saturated.

4.3. Comparisons with other SR methods

In order to further confirm the efficiency of the proposed LCAN, we quantitatively and visually compare our experimental results for upscaling factor ×2, ×3, and ×4 with the results of other state-of-the-art lightweight SR methods, including SRCNN (Dong et al., 2015), FSRCNN (Dong et al., 2016), VDSR (Kim et al., 2016a), DRCN (Kim et al., 2016b), DRRN (Tai et al., 2017), LapSRN (Lai et al., 2017), IDN (Hui et al., 2018), IMDN (Hui et al., 2019), PAN (Zhao et al., 2020), RFDN (Liu et al., 2020b), BSRN (Li et al., 2022), LKDN (Xie et al., 2023), MSID (Hu et al., 2023), DRSAN-32s (Park et al., 2021), LatticeNet-CL (Luo et al., 2022), SCN (Yang et al., 2023), and OSFFNet (Wang and Zhang, 2024). The self-ensemble strategy is used to further enhance our LCAN, and we denote it as LCAN+.

PSNR/SSIM results. Table 8 shows the quantitative evaluation results for ×2, ×3 and ×4 lightweight SR. For ×2 SR, with self-ensemble strategy, our LCAN+ has the best quantitative performance with the highest PSNR and SSIM values on all datasets. Besides, on five benchmark datasets, our LCAN achieves most of the highest PSNR and SSIM values in addition to having a minimum of 292K model parameters among these methods. For ×3 SR, with a minimum of 299K model parameters among them, our LCAN+ achieves the highest PSNR and SSIM values on five benchmark datasets, offering the finest quantitative performance, and our LCAN obtains most of the highest PSNR and SSIM. For ×4 SR, all the best quantitative performance, all the second PSNR results and all the second SSIM results are reached by our LCAN+, and our LCAN, respectively. Overall, the experimental results show that our LCAN can achieve superiority for lightweight SR, especially for large scaling factors (e.g., ×4) compared with previous lightweight state-of-the-art SR methods.

In addition, our LCAN and the heavyweight state-of-the-art SR methods are tested for ×4 SR on the benchmark datasets, and the results are compared in Table 7 in terms of model complexity and performance. As we can see, even when compared with the heavy SR networks, our model, which has at least 10 times less parameters, can produce competitive results. Specifically, our LCAN can respectively achieve 28.80 dB and 27.70 dB on Set14 and BSD100 datasets, which are equal to the corresponding results of the heavy network EDSR. Noted that the 0.309M parameters of our LCAN are far fewer than the 43.1 M parameters of EDSR. For Set5 and Urban100 datasets, the performance differences start to become large, reaching 0.03∼0.31 dB and 0.13∼0.56 dB, respectively. These results demonstrate that our LCAN can achieve a good trade-off between model complexity and performance.

Visual results. For ×2, ×3, and ×4 SR, Fig. 6 presents the visual comparison of different methods on specific images of Set14, BSD100, Manga109 and Urban100 datasets.

As for the images "img_024.png" and "img_092.png" for ×2 SR, there exists a lot of sharp contour information in the original HR images. We can find that the worst visual reconstructions are achieved by the early bicubic algorithm, while other advanced methods yield clearer details and more complete shapes. And the same trend goes for the quantitative results of PSNR and SSIM of these methods.

In terms of the images "barbara.png" and "img_098.png" for ×3 SR, CNN-based methods produce better visual results than the bicubic algorithm and obtain higher PSNR and SSIM values. Notably, in these methods, our LCAN produces more complete and natural contours and edges, and achieves higher PSNR and SSIM values, presenting superior reconstruction performance visually and quantitatively.

As for the images "148026.png" and "YumeiroCooking.png" for ×4 SR, the early bicubic algorithm still performs poorly with widespread blurring, aliasing artifacts, indicating its unstable trend for SR. Contrastively, the visual results of other CNN-based methods including IMDN, RFDN, BSRN and our LCAN, have been greatly improved, and they provide more contour information although they still have distorted shapes and blurring in some regions. The quantitative results of PSNR and SSIM also show that the reconstruction results of CNN-based methods are much better than those of the bicubic interpolation algorithm.

Overall, among these methods, our LCAN can yield clearer visual results and achieves the highest PSNR and SSIM for these test images, which show the effectiveness of the extracted discriminative feature with multi-scale information and more contextual information.

Table 4
Ablation studies of MBSConv on the Urban100 (Huang et al., 2015) dataset for ×4 SR. The input feature maps of MBSConv have 55 channels and the numbers in parentheses are the corresponding number of depth-wise convolutional kernels. The result of our adopted lightweight MBSConv with 1 × 1, 3 × 3, 5 × 5 kernels in our LCAN is underlined.

MBSConv	1 × 1 (18), 3 × 3 (18), 5 × 5 (19)	✓	✗	✗	✗	✗	✗	✗
	1 × 1 (13), 3 × 3 (13), 5 × 5 (13), 7 × 7 (16)	✗	✓	✗	✗	✗	✗	✗
	3 × 3 (18), 5 × 5 (18), 7 × 7 (19)	✗	✗	✓	✗	✗	✗	✗
BSConv	1 × 1 (55)	✗	✗	✗	✓	✗	✗	✗
	3 × 3 (55)	✗	✗	✗	✗	✓	✗	✗
	5 × 5 (55)	✗	✗	✗	✗	✗	✓	✗
	7 × 7 (55)	✗	✗	✗	✗	✗	✗	✓
Params/K	–	309	326	335	297	309	331	364
PSNR/dB	–	26.47	26.53	26.60	26.25	26.40	26.49	26.51

Table 5
Ablation studies of LCKA on the Manga109 (Matsui et al., 2017) dataset for ×4 SR. The TRFS denotes the theoretical receptive field size of single module. We adopt the LCKA with the strip size k as 9 in our LCAN. The best result is **highlighted**.

strip size k	5	7	9	5	7	9
dr = (k+1)/2	3	4	5	3	4	5
DWC, DWDC, PC	✓	✓	✓	✗	✗	✗
DWDC, DWC, PC	✗	✗	✗	✓	✓	✓
TRFS	17 × 17	31 × 31	49 × 49	17 × 17	31 × 31	49 × 49
PSNR/dB	30.76	30.91	**31.03**	30.71	30.85	30.99

Table 6
The further ablation studies of LCKA on the Manga109 (Matsui et al., 2017) dataset for ×4 SR. The strip size k is set to 9 in these modules. The best result is **highlighted**.

(H-DWC, H-DWDC, V-DWC, V-DWDC, PC)	✓	✗	✗
(H-DWC, H-DWDC, V-DWDC, V-DWC, PC)	✗	✓	✗
(H-DWDC, H-DWC, V-DWC, V-DWDC, PC)	✗	✗	✓
PSNR/dB	**31.03**	30.99	30.98

Table 7
Quantitative comparison of our LCAN with the heavy state-of-the-art SR methods on benchmark datasets for ×4 SR.

Method	Params/M	PSNR/dB			
		Set5	Set14	BSD100	Urban100
EDSR	43.1	32.46	28.80	27.70	26.64
RDN	22.3	32.47	28.80	27.72	26.60
RCAN	15.6	32.63	28.87	27.77	26.82
DRN	9.8	32.74	28.98	27.83	27.03
ERAN	8.02	32.66	28.92	27.79	26.86
LCAN	0.309	32.43	28.80	27.70	26.47

4.4. Experiments on real-world photos

To validate the model performance in practice, we also run test experiments on real LR images that degenerate in an unknown manner. Since there are none ground-truth HR images for these real LR images, we only provide visual results for comparison. We select images "pattern.png" and "Historical_006" as test images and compare our LCAN with five SR methods, including early bicubic algorithm (Zhang and Wu, 2006), SRCNN (Dong et al., 2015), RFDN (Liu et al., 2020b), LKDN (Xie et al., 2023), and MSID (Hu et al., 2023). As shown in Figs. 7 and 8, our LCAN yields natural SR visual results with clearer contour information and sharper details, while early methods, i.e., bicubic and SRCNN, yield distorted and unstable results with a lot of compression artifacts, and other methods including RFDN, LKDN, and MSID produce better SR visual results than the early methods. Overall, our LCAN can yield better or comparable visual performance than other methods, confirming its effectiveness and robustness when applied to real-world images.

4.5. Model complexity analysis

Now, according to the recorded results, we analyze the model complexity and reconstruction performance of different methods on Urban100 dataset for ×2 SR. As widely used evaluation metrics, model parameters and multi-adds are used to quantitatively present the model complexity, and the PSNR is utilized to evaluate the reconstruction quality of visual outputs. All quantitative results of these SR methods, including LapSRN (Lai et al., 2017), IMDN (Hui et al., 2019), RFDN (Liu et al., 2020b), LKDN (Xie et al., 2023), MSID (Hu et al., 2023), LatticeNet-CL (Luo et al., 2022), and our LCAN, are listed in Table 9. We can see that our LCAN achieves the highest PSNR with the fewest model parameters, which demonstrates the efficiency of our network. Noted that our LCAN with only 291.6K parameters is the most lightweight network while achieving the highest PSNR 32.62 dB on Urban100 dataset for ×2 SR.

Additionally, the multi-adds defined in Ahn et al. (2018) denotes the number of multiply accumulate operations, and it is another quantitative evaluation metric for model complexity. To fairly compare our model with other state-of-the-art SR networks, following (Shi et al., 2023; Xie et al., 2023), we also assume the SR output size to 1280 × 720 to calculate the model multi-adds. To gain a better understanding of model complexity and performance, we compare our LCAN with other methods on Urban100 for ×2 SR in terms of PSNR and Multi-Adds. Table 9 shows the detailed comparison results. As we can see, our LCAN has only 68.37 G multi-adds, which are 0.73 G fewer than 69.1 G multi-adds of LKDN, and it can achieve the highest PSNR 32.62 dB, outperforming LKDN by 0.09 dB and surpassing other SR methods by a large gap. The comparison results show that our proposed LCAN is effective and efficient.

4.6. Inference time

With fewer parameters and activities, our LCAN could obtain superior performance, as shown in Table 8. To further evaluate model efficiency, we compute the practical inference times of different advanced SR models. Using official codes of these compared methods, we compute their average inference times on one hundred 1280 × 720 super-resolved RGB images under the same experimental environment as ours. The PyTorch framework's torch.cuda.Event is used to compute inference times of these methods. For ×4 SR, Table 10 presents the

Table 8
Quantitative outcomes of our method and other state-of-the-art lightweight SR methods on five benchmark datasets. The methods with † and ‡ are retrained based on their official codes and tested on their open-source models, respectively, and for the rest methods without both, we also directly adopt the results in their original papers as in the previous methods (Li et al., 2022; Yang et al., 2023; Wang and Zhang, 2024). We respectively **highlight** and underline the best and the second-best results.

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
DRRN ‡		298K	37.74/0.9591	33.23/0.9136	32.05/0.8973	31.23/0.9188	37.88/0.9749
LapSRN ‡	×2	251K	37.52/0.9591	32.99/0.9124	31.80/0.8952	30.41/0.9103	37.27/0.9740
IDN ‡		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
LKDN ‡		304K	38.12/**0.9611**	33.90/0.9202	32.27/0.9010	32.53/0.9322	39.19/0.9784
MSID ‡		375K	38.10/0.9609	33.84/0.9198	32.29/0.9012	32.57/0.9326	39.23/0.9783
DRSAN-32s		370K	37.99/0.9606	33.57/0.9177	32.16/0.8999	32.10/0.9279	–/–
LatticeNet-CL		756K	38.09/0.9608	33.70/0.9188	32.21/0.9000	32.29/0.9291	–/–
SCN		688K	38.10/0.9608	33.82/0.9200	32.28/0.9010	32.57/0.9324	39.12/0.9776
LCAN (ours)		292K	38.12/0.9610	33.90/0.9202	32.30/0.9013	32.62/0.9331	39.28/0.9783
LCAN+ (ours)		292K	**38.14/0.9611**	**33.94/0.9203**	**32.31/0.9014**	**32.78/0.9342**	**39.38/0.9785**
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
DRRN ‡		298K	34.03/0.9244	29.96/0.8349	28.95/0.8004	27.53/0.8378	32.71/0.9379
LapSRN ‡	×3	502K	33.81/0.9220	29.79/0.8325	28.82/0.7980	27.07/0.8275	32.21/0.9350
IDN ‡		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
LKDN ‡		311K	34.54/0.9285	30.52/0.8455	29.21/0.8078	28.50/0.8601	34.08/0.9475
MSID ‡		383K	34.54/0.9283	30.51/0.8456	29.22/0.8083	28.53/0.8603	34.14/0.9477
DRSAN-32s		410K	34.41/0.9272	30.27/0.8413	29.08/0.8056	28.19/0.8529	–/–
LatticeNet-CL		765K	34.46/0.9275	30.37/0.8422	29.12/0.8054	28.23/0.8525	–/–
SCN		697K	34.52/0.9281	30.44/0.8443	29.15/0.8070	28.45/0.8582	33.96/0.9467
LCAN (ours)		299K	34.58/0.9286	30.53/0.8456	29.22/0.8082	28.57/0.8613	34.18/0.9478
LCAN+ (ours)		299K	**34.63/0.9289**	**30.57/0.8461**	**29.25/0.8087**	**28.68/0.8628**	**34.32/0.9485**
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
DRRN ‡		298K	31.68/0.8888	28.21/0.7720	27.38/0.7284	25.44/0.7638	29.45/0.8946
LapSRN ‡	×4	502K	31.54/0.8852	28.09/0.7700	27.32/0.7275	25.21/0.7562	29.09/0.8900
IDN ‡		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
LKDN ‡		322K	32.39/0.8979	28.79/0.7859	27.69/0.7402	26.42/0.7965	30.97/0.9140
MSID ‡		393K	32.40/0.8973	28.79/0.7858	27.69/0.7401	26.43/0.7966	30.94/0.9132
DRSAN-32s		410K	32.15/0.8935	28.54/0.7813	27.54/0.7364	26.06/0.7858	–/–
LatticeNet-CL		777K	32.30/0.8958	28.65/0.7822	27.59/0.7365	26.19/0.7855	–/–
SCN		709K	32.34/0.8967	28.71/0.7834	27.62/0.7381	26.31/0.7926	30.76/0.9112
OSFFNet		537K	32.39/0.8976	28.75/0.7852	27.66/0.7393	26.36/0.7950	30.84/0.9125
LCAN (ours)		309K	32.43/0.8979	28.80/0.7860	27.70/0.7405	26.47/0.7980	31.03/0.9145
LCAN+ (ours)		309K	**32.49/0.8985**	**28.84/0.7867**	**27.73/0.7412**	**26.54/0.7994**	**31.17/0.9155**

Table 9
Computation and parameters comparison (×2 Urban100), The best results are **highlighted**.

Metric	IMDN	RFDN	LKDN	DRSAN-32s	LatticeNet-CL	LCAN (ours)
Params/K	694	535	304	370	756	**291.6**
Multi-Adds/G	158.8	95	69.1	85.5	169.5	**68.37**
PSNR/dB	32.17	32.12	32.53	32.1	32.29	**32.62**

inference times for the proposed LCAN and the state-of-the-art SR methods. Theoretically, depth-wise convolution can lessen the standard convolution's computational complexity and number of parameters (Haase and Amthor, 2020; Chollet, 2017).

However, the high memory access cost (MAC) to floating-point operations (FLOPs) of depth-wise convolution means that the acceleration performance of GPU is currently unable to reach the theoretical value. Therefore, as shown in Table 10, our method offers no significant

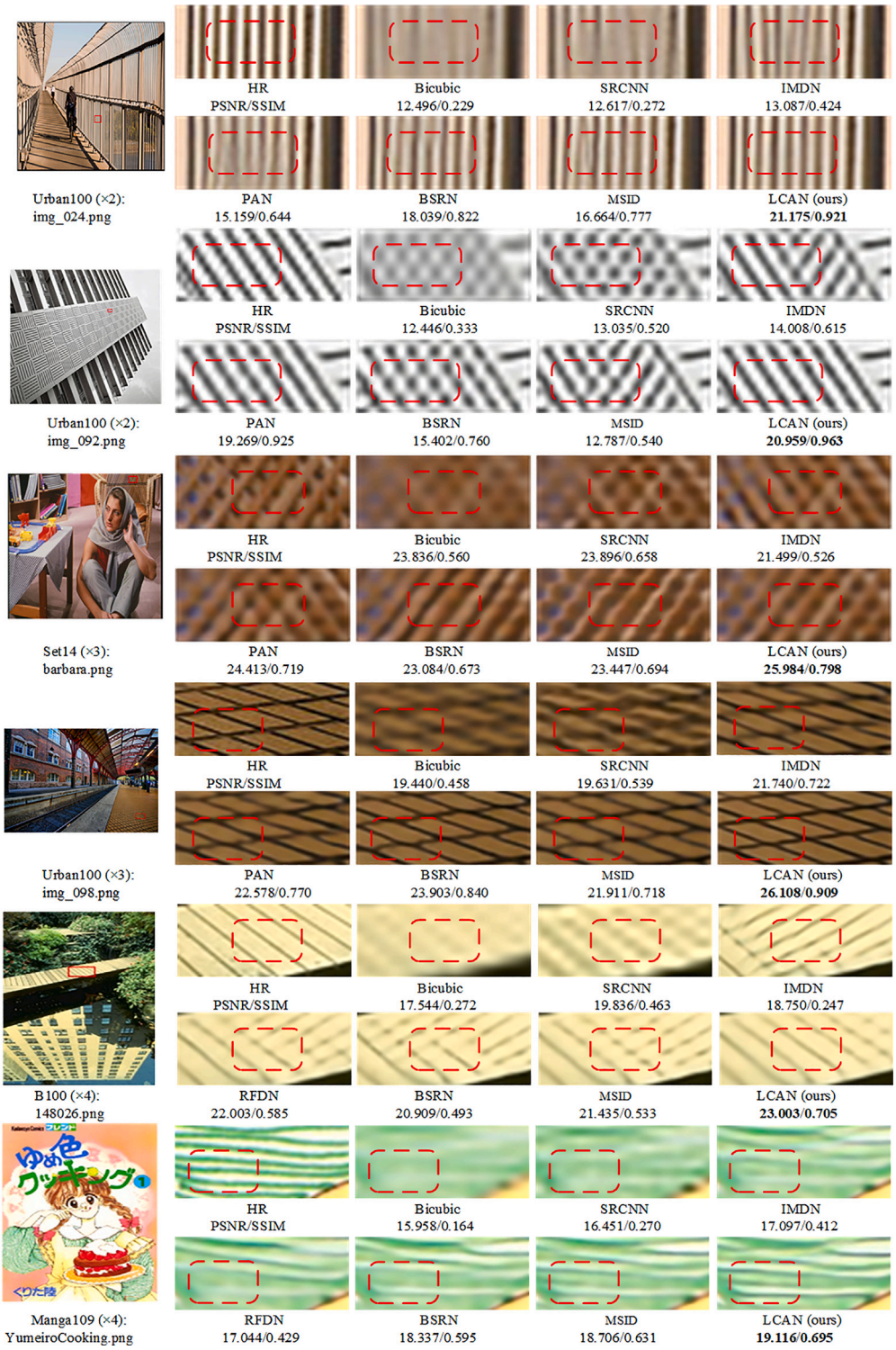


Fig. 6. Visual comparisons for ×2, ×3, and ×4 SR on the Set14, B100, Urban100 and Manga109 datasets with the BI degradation. We **highlight** the best results.

Table 10
The performance on Set5 dataset and average inference times in millisecond for various advanced methods on 100 SR RGB pictures with 1280 × 720 for ×4 SR. The best results are **highlighted**.

Method	NGswin	IMDN	RFDN	PAN	BSRN	LKDN	MSID	LCAN (ours)
Time/ms	157.36	8.93	**7.61**	17.26	17.93	20.97	18.77	23.95
PSNR/dB	32.33	32.21	32.24	32.13	32.35	32.39	32.40	**32.43**

advantages in terms of practical inference time over other lightweight CNN-based approaches, even if our LCAN has fewer parameters and FLOPs in theory. It is worth noting that our method infers faster than transformer-based method, NGswin Choi et al. (2023). Further research

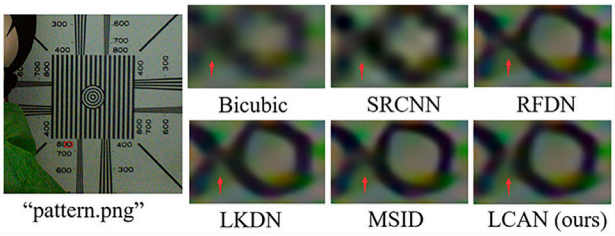


Fig. 7. Comparison of real image "pattern.png" for ×4 SR.

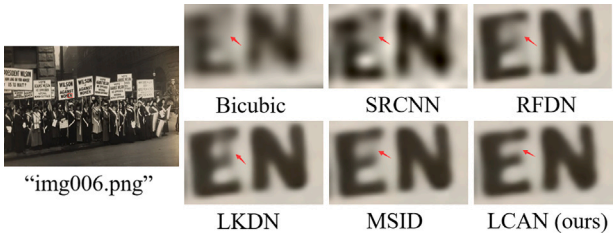


Fig. 8. Comparison of historical image "img006.png" for ×4 SR .

is required to improve the implementation efficiency of depth-wise convolution in order to speed up its feed-forward process, which would further ease the development of lightweight networks.

5. Conclusions

We propose a Large Coordinate Attention Network (LCAN) which is extremely lightweight for efficient image SR. Specifically, we propose multi-scale blueprint separable convolutions (MBSConv) to optimize intra-kernel correlations for discriminative feature with multi-scale information. In addition, we propose a novel large coordinate kernel attention (LCKA) module which enables the adjacent direct interaction of local information and long-distance dependencies in horizontal and vertical directions, respectively. Besides, the LCKA allows for the direct use of extremely large kernels in the depth-wise convolutional layers to capture more contextual information, which helps to significantly improve the reconstruction performance, and it incurs lower computational complexity and memory footprints. Extensive experiments quantitatively and visually demonstrate the superiority, efficiency and robustness of our proposed LCAN for lightweight SR, on the datasets created via BI degradation and on the real-world photos. For future works, the inference speed of our LCAN needs to be further improved, and we hope our MBSConv and LCKA can succeed on other low-level vision tasks (e.g., denoising and deblurring) and the high-level vision tasks (e.g., image classification and object detection).

CRediT authorship contribution statement

Fangwei Hao: Writing – original draft, Software, Methodology, Data curation, Conceptualization. **Jiesheng Wu:** Visualization, Supervision. **Haotian Lu:** Validation, Software. **Ji Du:** Software, Investigation. **Jing Xu:** Writing – review & editing. **Xiaoxuan Xu:** Supervision,Project administration.

Declaration of competing interest

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

Data availability

Data will be made available on request.

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. ECCV, pp. 252–268.

Arbelaez, P., Maire, M., Fowlkes, C., Malik, J., 2010. Contour detection and hierarchical image segmentation. 33, (5), IEEE, pp. 898–916,

Berardini, D., Migliorelli, L., Galdelli, A., Marín-Jiménez, M.J., 2025. Edge artificial intelligence and super-resolution for enhanced weapon detection in video surveillance. Eng. Appl. Artif. Intell. 140, 109684.

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

Cheng, G., Matsune, A., Du, H., Liu, X., Zhan, S., 2022. Exploring more diverse network architectures for single image super-resolution. Knowl.-Based Syst. 235, 107648.

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. ICPR, IEEE, pp. 59–64.

Dong, C., Loy, C.C., He, K., Tang, X., 2015. Image super-resolution using deep convolutional networks. IEEE Trans. Pattern Anal. Mach. Intell. 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. Springer, pp. 391–407.

Feng, C.M., Yan, Y., Yu, K., Xu, Y., Fu, H., Yang, J., Shao, L., 2024. Exploring separable attention for multi-contrast MR image super-resolution. IEEE Trans. Neural Netw. Learn. Syst..

Gao, G., Li, W., Li, J., Wu, F., Lu, H., Yu, Y., 2022. Feature distillation interaction weighting network for lightweight image super-resolution. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 36, (1), pp. 661–669.

Guo, M.H., Lu, C.Z., Liu, Z.N., Cheng, M.M., Hu, S.M., 2023. Visual attention network. Comput. Vis. Media 9 (4), 733–752.

Haase, D., Amthor, M., 2020. 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., Wu, J., Liang, W., Xu, J., Li, P., 2024. Lightweight blueprint residual network for single image super-resolution. Expert Syst. Appl. 250, 123954.

Hao, F., Zhang, T., Zhao, L., Tang, Y., 2022. Efficient residual attention network for single image super-resolution. Appl. Intell. 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.

Hu, Y., Huang, Y., Zhang, K., 2023. Multi-scale information distillation network for efficient image super-resolution. Knowl.-Based Syst. 275, 110718.

Hu, Y., Li, J., Huang, Y., Gao, X., 2019. Channel-wise and spatial feature modulation network for single image super-resolution. IEEE Trans. Circuits Syst. Video Technol. 30 (11), 3911–3927.

Hu, J., Shen, L., Sun, G., 2018. Squeeze-and-excitation networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 7132–7141.

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.

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.

Kong, D., Gu, L., Li, X., Gao, F., 2024. Multi-scale residual dense network for the super-resolution of remote sensing images. IEEE Trans. Geosci. Remote Sens..

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 Trans. Pattern Anal. Mach. Intell. 41 (11), 2599–2613.

Lau, K.W., Po, L.M., Rehman, Y.A.U., 2024. Large separable kernel attention: Rethinking the large kernel attention design in cnn. Expert Syst. Appl. 236, 121352.

Li, Z., Liu, Y., Chen, X., Cai, H., Gu, J., Qiao, Y., Dong, C., 2022. Blueprint separable residual network for efficient image super-resolution. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 833–843.

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.

Liu, H., Cao, F., Wen, C., Zhang, Q., 2020a. Lightweight multi-scale residual networks with attention for image super-resolution. Knowl.-Based Syst. 203, 106103.

Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., Guo, B., 2021a. Swin transformer: Hierarchical vision transformer using shifted windows. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 10012–10022.

Liu, J., Tang, J., Wu, G., 2020b. 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. Springer, pp. 41–55.

Liu, N., Zhang, N., Shao, L., Han, J., 2021. Learning selective mutual attention and contrast for RGB-D saliency detection. IEEE Trans. Pattern Anal. Mach. Intell. 44 (12), 9026–9042.

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

Lu, Z., He, S., Li, D., Song, Y.-Z., Xiang, T., 2023. Prediction calibration for generalized few-shot semantic segmentation. IEEE Trans. Image Process. 32, 3311–3323.

Luo, X., Qu, Y., Xie, Y., Zhang, Y., Li, C., Fu, Y., 2022. Lattice network for lightweight image restoration. IEEE Trans. Pattern Anal. Mach. Intell. 45 (4), 4826–4842.

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

Park, K., Soh, J.W., Cho, N.I., 2021. A dynamic residual self-attention network for lightweight single image super-resolution. IEEE Trans. Multimed. 25, 907–918.

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.

Shi, H., Hayat, M., Cai, J., 2023. Transformer scale gate for semantic segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 3051–3060.

Sun, S., Yue, X., Zhao, H., Torr, P.H., Bai, S., 2022. Patch-based separable transformer for visual recognition. IEEE Trans. Pattern Anal. Mach. Intell. 45 (7), 9241–9247.

Symeonidis, C., Mademlis, I., Pitas, I., Nikolaidis, N., 2023. Neural attention-driven non-maximum suppression for person detection. IEEE Trans. Image Process. 32, 2454–2467.

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 Trans. Image Process. 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.

Wang, Y., Zhang, T., 2024. Osffnet: Omni-stage feature fusion network for lightweight image super-resolution. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 38, (6), pp. 5660–5668.

Xie, C., Zhang, X., Li, L., Meng, H., Zhang, T., Li, T., Zhao, X., 2023. Large kernel distillation network for efficient single image super-resolution. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 1283–1292.

Xie, X., Zhou, P., Li, H., Lin, Z., Yan, S., 2024. Adan: Adaptive nesterov momentum algorithm for faster optimizing deep models. IEEE Trans. Pattern Anal. Mach. Intell..

Yang, H., Yang, X., Liu, K., Jeon, G., Zhu, C., 2023. SCN: Self-calibration network for fast and accurate image super-resolution. Expert Syst. Appl. 226, 120159.

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. Springer, pp. 711–730.

Zhang, K., Gu, S., Timofte, R., Hui, Z., Wang, X., Gao, X., Xiong, D., Liu, S., Gang, R., Nan, N., et al., 2019. Aim 2019 challenge on constrained super-resolution: Methods and results. In: 2019 IEEE/CVF International Conference on Computer Vision Workshop. ICCVW, IEEE, pp. 3565–3574.

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

Zhang, L., Ma, K., 2023. Structured knowledge distillation for accurate and efficient object detection. IEEE Trans. Pattern Anal. Mach. Intell. 45 (12), 15706–15724.

Zhang, F., Panahi, A., Gao, G., 2023. FsaNet: Frequency self-attention for semantic segmentation. IEEE Trans. Image Process. 32, 4757–4772.

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

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

Zhang, Q.L., Yang, Y.B., 2021. Sa-net: Shuffle attention for deep convolutional neural networks. In: ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing. ICASSP, IEEE, pp. 2235–2239.

Zhang, K., Zuo, W., Zhang, L., 2018c. 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. Springer, pp. 56–72.

Zhou, L., Cai, H., Gu, J., Li, Z., Liu, Y., Chen, X., Qiao, Y., Dong, C., 2022. Efficient image super-resolution using vast-receptive-field attention. In: European Conference on Computer Vision. Springer, pp. 256–272.