



ARTICLE

A Lightweight Super-Resolution Network for Infrared Images Based on an Adaptive Attention Mechanism

Mengke Tang¹, Yong Gan^{2,*}, Yifan Zhang¹ and Xinxin Gan³

¹School of Computer Science and Technology, Zhengzhou University of Light Industry, Zhengzhou, 450001, China

²School of Information Engineering, Zhengzhou University of Technology, Zhengzhou, 450044, China

³Digital and Intelligent Engineering Design Institute, SIPPR Engineering Group Co., Ltd., Zhengzhou, 450007, China

*Corresponding Author: Yong Gan. Email: ganyong@zzuli.edu.cn

Received: 18 February 2025; Accepted: 27 April 2025; Published: 03 July 2025

ABSTRACT: Infrared imaging technology has been widely adopted in various fields, such as military reconnaissance, medical diagnosis, and security monitoring, due to its excellent ability to penetrate smoke and fog. However, the prevalent low resolution of infrared images severely limits the accurate interpretation of their contents. In addition, deploying super-resolution models on resource-constrained devices faces significant challenges. To address these issues, this study proposes a lightweight super-resolution network for infrared images based on an adaptive attention mechanism. The network's dynamic weighting module automatically adjusts the weights of the attention and non-attention branch outputs based on the network's characteristics at different levels. Among them, the attention branch is further subdivided into pixel attention and brightness-texture attention, which are specialized for extracting the most informative features in infrared images. Meanwhile, the non-attention branch supplements the extraction of those neglected features to enhance the comprehensiveness of the features. Through ablation experiments, we verify the effectiveness of the proposed module. Finally, through experiments on two datasets, FLIR and Thermal101, qualitative and quantitative results demonstrate that the model can effectively recover high-frequency details of infrared images and significantly improve image resolution. In detail, compared with the suboptimal method, we have reduced the number of parameters by 30% and improved the model performance. When the scale factor is 2, the peak signal-to-noise ratio of the test datasets FLIR and Thermal101 is improved by 0.09 and 0.15 dB, respectively. When the scale factor is 4, it is improved by 0.05 and 0.09 dB, respectively. In addition, due to the lightweight design of the network structure, it has a low computational cost. It is suitable for deployment on edge devices, thus effectively enhancing the sensing performance of infrared imaging devices.

KEYWORDS: Infrared image; super-resolution; convolutional neural network; attention mechanism; dynamic network

1 Introduction

Infrared images can offer significant thermal data in circumstances where visual images are unsuitable, such as in darkness, fog, or smoke, and have important applications in various fields, such as medical, biochemical engineering, and vision tasks. However, a critical challenge in infrared imaging is the lack of detail resolution, which severely limits image visualization and target recognition performance. The lower spatial resolution of infrared sensors compared to Red, Green, and Blue (RGB) cameras is a major hardware limitation that contributes to this issue. Therefore, algorithmic solutions are essential to bridge this gap. To better understand the challenges and potential solutions, it is important to first examine the



limitations of infrared imaging systems at the hardware level. In a classical infrared imaging system, the following basic components are included: lens, sensor, processor, etc. [1,2]. Although boosting the imaging system's performance can improve infrared image quality, such enhancements face significant cost barriers and insurmountable physical constraints. Algorithmic techniques are essential to improve the resolution of infrared images due to hardware restrictions. In recent years, a technique referred to as Single Image Super Resolution (SISR) has been extensively advanced to improve the resolution of low-resolution images [3]. The objective of SISR is to reconstruct the associated high-resolution image from the observed degraded low-resolution image. Without changing the hardware imaging system, SISR techniques have significantly demonstrated the potential to improve image quality, especially in resource-limited environments [4,5].

Image super-resolution methods can be classified into two primary categories: conventional approaches and those based on deep learning. The traditional approaches can be further classified into three categories: frequency domain techniques, dictionary-based procedures, and assorted other ways. These approaches utilize prior knowledge to limit the solution space and derive the solution with minimal distance to the actual HR [6–8]. Nevertheless, weak mathematical analysis limits the proposed algorithm, and these conventional methods exhibit a lack of robustness and fail to adapt to the intricate and dynamic nature of infrared images. With the development of deep learning, it has become popular to use a Convolutional Neural Network (CNN) to build an end-to-end model for image super-resolution. Neural networks can learn nonlinear mappings without instructions with their powerful fitting ability [9–11] and have shown excellent performance in SISR, freeing people from the task of finding priors. CNN-based SISR methods convert low-resolution images to high-resolution ones through an end-to-end approach utilizing the training dataset. Leveraging the robust nonlinear fitting and automatic feature extraction abilities of convolutional neural networks, along with the advent of specialized hardware like neural processing units (NPU), the efficacy of image super-resolution networks that have been trained on extensive datasets far surpasses that of conventional techniques. Super-Resolution Convolutional Neural Network (SRCNN) [9] is the inaugural CNN-based SISR model, demonstrating markedly enhanced performance compared to interpolation and sparse coding approaches. Leveraging the robust nonlinear fitting capabilities of CNNs, numerous CNN-based RGB sensor single-image super-resolution models have been consistently introduced [12]. Nonetheless, the majority of current deep learning approaches for enhancing infrared images through super-resolution are adapted from techniques developed for CNN-based visible light image enhancement. This adaptation often overlooks the unique characteristics of infrared images and RGB images, such as their lower signal-to-noise ratios, diminished contrast, and blurred visual qualities, all of which can significantly impact the model's effectiveness. Secondly, current CNN-based infrared image super-resolution models mainly suffer from low efficiency in the sequential stacking of attention modules, complex and bulky network structures, and non-compact attention module designs.

Given these challenges, this study aims to address the following research questions:

- (1) How can we develop a CNN-based super-resolution model specifically tailored for infrared images, considering their unique characteristics such as lower signal-to-noise ratios, diminished contrast, and blurred visual qualities?
- (2) How can we design a lightweight super-resolution network that balances efficiency and performance while effectively enhancing the resolution and contrast of infrared images?
- (3) How can we incorporate adaptive attention mechanisms into the network to better capture detailed edges and brightness-texture features of infrared images, thereby improving the overall quality of super-resolution results?

To address these problems, we present the Lightweight Dynamic Super-Resolution Network (LDynSR), a lightweight network for infrared image super-resolution that leverages an adaptive attention mechanism

to effectively enhance resolution. Firstly, considering the challenges of low contrast and blurred textures in infrared images, we propose a brightness-texture attention mechanism to prevent detail loss caused by low contrast while restoring high-frequency details. Additionally, pixel attention performs pixel-level refinement to reduce noise and mitigate local detail degradation in infrared images. Finally, a dynamic weighting module is introduced to adaptively adjust the weights of attention and non-attention branches, improving information utilization. While maintaining a lightweight architecture, the model effectively enhances the contrast of infrared images, refines image details, and reconstructs high-resolution, high-quality infrared images.

This paper's primary contributions are as follows:

- A CNN-based adaptive attention mechanism network is proposed to perform the task of super-resolution of infrared images. The proposed model balances efficiency and performance by better learning the mapping relationship between LR images and HR images for infrared images, enhancing feature expression, and better reconstructing high-resolution infrared images while maintaining light weight.
- A lightweight dynamic weighting module, Dynamic Weighting Module (DAM), is designed to be applied to different depths of the attention mechanism. Different stages of the DAM dynamically adjust the weights of attention and non-attention paths. The proposed attention branches include a pixel attention block and a brightness-texture attention block, which are used to focus on the detailed edges and brightness texture features of the infrared image, respectively.
- The quantitative and qualitative studies conducted on various datasets demonstrate that the proposed model achieves superior performance metrics in optimizing the balance between the number of model parameters and the restoration of high-frequency details in infrared images.

2 Related Work

2.1 Visible Light Image Super-Resolution Based on CNN

Given CNN's exceptional precision in image identification [13], CNN-based SISR techniques have garnered considerable interest. Dong et al. [9] proposed SRCNN, the inaugural model to effectively utilize CNN for picture super-resolution. It proposed a method to directly establish the correspondence between low-resolution and high-resolution images, avoiding the manual feature design and complex optimization process in traditional methods. SRCNN contains only three convolutional layers and is capable of learning the nonlinear mapping between low-resolution and high-resolution images in an end-to-end manner, surpassing conventional single-image super-resolution techniques. Subsequently, Fast Super-Resolution Convolutional Neural Network (FSRCNN) [10] further enhanced SRCNN by designing a more efficient network structure. FSRCNN significantly improved the processing speed of super-resolution tasks while also improving performance. However, although FSRCNN is fast, it may still not be able to recover enough image details when processing some extremely low-resolution images, especially at high magnifications, where detail loss is more obvious. To enhance the efficacy of deep learning models, the Very Deep Super-Resolution network (VDSR) [11] employs 20 convolutional layers to more effectively capture visual details and high-order characteristics. In addition, VDSR also introduces residual learning to address issues like gradient explosion and convergence difficulties in deep networks, hence enhancing the efficacy of SISR. Enhanced Deep Super-Resolution Network (EDSR) [14] comprises 65 convolutional layers. EDSR enhances model performance by eliminating the batch normalization layer, which boosts the model's detail recovery capability, diminishes noise in parameter updates during training, and increases training stability. Residual Channel-Attention Network (RCAN) [15] is a model featuring over 400 convolutional layers and was the first to propose and widely apply the channel attention mechanism in image super-resolution. Dynamically

weighting the importance of each channel effectively enhances the model's capacity to restore high-resolution image information.

However, as network depth increases, the requirement for processing resources and memory also increases accordingly. Particularly when handling high-resolution images, where the requirements for memory and computation become especially pronounced, complicating application on resource-constrained devices. Recognizing these problems, many researchers are now focusing on developing lightweight models for SISR. Deeply-Recursive Convolutional Network (DRCN) [16] expands the depth of the convolutional layer by deep recursion instead of simply adding new convolutional layers. This method can expand the receptive field without adding additional parameters, thereby using greater contextual information to restore high-frequency details of the image, solving the problem of increasing the network depth in traditional methods, leading to a sharp increase in the number of parameters. However, since the recursive depth of DRCN can reach 16 layers, gradient vanishing and gradient exploding problems are prone to occur during training. These problems make it difficult for the network to converge, and the training process becomes very unstable, requiring a lot of parameter adjustment and optimization techniques to solve. Cascading Residual Network (CARN) [17] reduces the parameters and computation of the model by using efficient residual blocks (residual-E) and recursive network architecture. The residual E block significantly reduces the computational effort and the number of parameters of the convolution operation by using group convolution instead of the standard convolution operation. Furthermore, in Cascading Residual Network with Shared Parameters (CARN-M), the parameters of the cascade blocks are shared, avoiding the need to assign separate parameters to each cascade block, further significantly reducing the total number of parameters in the model. Although the cascade structure of CARN can improve the performance of the model, this structure also makes the convergence speed during training slower. Due to the complexity of the cascade modules, the model requires more training time to achieve better performance, which increases the training cost and time. Information Multi-Distillation Network (IMDN) [18] uses Information Multi-Distillation Blocks (IMDB) to achieve a lightweight design. IMDB extracts features step by step through Progressive Refinement Modules (PRMs), where features are divided into retained portions and portions to be further processed in each step. The retained portions are considered refinement features, thus reducing unnecessary computations. However, this gradual refinement may not be enough to capture the deep features of the image in some cases. Especially when dealing with complex scenes and high-resolution images, deeper feature extraction may be required to achieve better super-resolution effects. The Residual Feature Distillation Network (RFDN) [19] enhances the channel division of IMDN, resulting in increased efficiency and reduced weight. However, this enhancement may come at the cost of reduced feature extraction depth. Pixel Attention Network (PAN) [20] uses lightweight self-calibrated convolution [21] as a basic building block to directly learn the important weights of features at the pixel level, strengthen attention to key pixels, and improve the ability to restore image texture and details. However, this approach may struggle with complex backgrounds or low-contrast images, as it tends to focus on prominent regions and neglects less noticeable background details, which can result in suboptimal restoration performance in certain cases.

2.2 Infrared Image Super-Resolution Based on CNN

The remarkable advancement of CNN-based RGB image super-resolution has likewise facilitated the progression of infrared image super-resolution. Motivated by SRCNN, a thermal image enhancement network architecture called TEN [22] was proposed. It is a relatively shallow neural network containing only four convolutional layers. Nevertheless, because of the exorbitant expense of HR thermal detectors at the time and the difficulty in collecting sufficient LR-HR infrared images for training, TEN utilized RGB images as the training dataset. However, Rivadeneira et al. [23] demonstrated that TEN trained on RGB images as

its dataset, which limited the infrared image SR performance. The computationally efficient Thermal Image Super-Resolution Network (TherISuRNet) [24] employs a progressive upsampling technique to achieve thermal image super-resolution efficiently, incrementally enhancing the image resolution to meet the desired resolution. Despite its efficiency, this approach may accumulate errors at each upsampling stage, potentially degrading the final image quality. Channel split convolutional neural network (ChaSNet) [25] uses channel segmentation technology to reduce redundant features in the network. This method reduces the information load and optimizes the performance of deep models by distinguishing features within the channel dimension. However, this technique may also inadvertently discard some useful information, potentially leading to a slight drop in image quality or detail recovery, especially in complex scenes where diverse features are critical for accurate reconstruction. Zou et al. [26] suggested a CNN architecture consisting of convolutional and deconvolutional layers, where the two halves of the network are connected through skip connections to transfer image information and mitigate the gradient vanishing problem. While this improved super-resolution performance, the architecture may suffer from overfitting when model complexity is too high for the available dataset, as skip connections could lead to excessive feature propagation, amplifying noise, and reducing the model's ability to generalize. Du et al. [27] introduced a hybrid convolution consisting of standard convolution and dilated convolution. This novel technology expands the receptive field without altering the dimensions of the feature map and eradicates blind spots. Additionally, a recursive fusion method is employed to process feature outputs of varying scales more smoothly to achieve better reconstruction effects. However, the combination of standard and dilated convolutions, along with recursive fusion, might increase the model's complexity and computational cost. This could result in slower training and inference times, which may be a drawback for real-time applications or when processing large volumes of data. The stacked Multiscale Feature Distillation Residual Blocks (MFDRB) [28] are constructed from three residual distillation blocks, designed for mobile infrared imaging systems to perform channel feature distillation operations and extract multiscale features. Despite their effectiveness, the reliance on multiple distillation blocks may increase the computational burden, leading to longer processing times and higher memory consumption, challenges for real-time applications on mobile devices with limited resources. Given the limitations of existing models, such as high computational cost, overfitting, and image quality degradation, there is a clear need for a more efficient solution. Many current approaches either struggle with preserving details or are too complex for resource-constrained environments. Our proposed model addresses these issues by providing a lightweight architecture with fewer parameters, balancing performance and efficiency. Unlike previous CNN-based infrared super-resolution models, our proposed LDynSR integrates a lightweight architecture, a dynamic attention module (DynA), and specialized attention mechanisms to enhance performance. The lightweight design reduces computational costs, making it suitable for resource-constrained devices. Unlike models that rely on fixed attention mechanisms, the dynamic attention module we propose adaptively adjusts the contributions of the attention and non-attention branches, thereby increasing the model's flexibility and focus on critical details. Additionally, the pixel attention (PA) and brightness-texture attention (BTA) mechanisms specifically address challenges in infrared images, such as low contrast and texture loss, achieving superior resolution and detail recovery. Collectively, these innovations achieve strong performance while maintaining efficiency.

2.3 Attention Mechanism

Attention is a core part of the human cognitive system, which allows us to filter and focus on specific information among various sensory inputs. When processing substantial information, we can concentrate on significant things without being overwhelmed by minor things. Mimicking this ability of humans, the

attention mechanism in neural networks implements attention to specific information, enabling the model to selectively highlight important features while suppressing less relevant ones.

The squeeze-excitation (SE) [29] network is constructed with the SE block as its central component. The SE block mainly includes two key steps: squeeze and excitation. The squeeze operation aims to compress global spatial information into channel descriptors. This step is implemented by global average pooling (GAP), which aggregates the information of each feature map in the spatial dimension (i.e., height and width) to obtain a channel-level statistic. The purpose of the excitation operation is to dynamically adjust the feature response of each channel using the channel descriptor generated by the squeeze operation. This step involves a self-gating mechanism that assesses the significance of each channel according to the inter-channel dependencies. Subsequently, Guha Roy et al. [30] initially proposed the squeeze and excitation (sSE) module in the spatial dimension and the squeeze and excitation (scSE) module in the spatial and channel dimensions. These modules allow the model to focus more on significant regions and channels in the image, thereby enhancing its accuracy. The Convolutional Block Attention Module (CBAM) [31] attention mechanism module contains two sequential submodules, namely, the channel attention module and the spatial attention module. This dual attention mechanism enables CBAM to improve the expressiveness of feature maps, raise the model's capacity to recognize and find targets, and ultimately improve the performance of CNN in diverse visual tasks. The Dual Attention Network (DANet) [32] incorporates the position attention module (PAM) and channel attention module (CAM). The PAM selectively collects features from each position by assigning weights to the features of all positions, while the CAM highlights interdependent channel maps by integrating relevant information across all channel maps. Efficient Channel Attention (ECA) [33] generates channel attention via rapid one-dimensional convolution, and the kernel size of the one-dimensional convolution can be adaptively determined according to the channel dimension. The Channel-Wise and Spatial Feature Modulation (CSFM) [34] network uses channel attention (CA) and spatial attention (PA) mechanisms to calibrate feature maps, respectively, and then combines the results of the two types of attention.

3 Proposed Method

3.1 Network Architecture

As shown in Fig. 1, our proposed network architecture LDynSR has three modules, which are composed of a 3×3 convolutional layer, a deep feature extraction module (DFEB), and a feature reconstruction module (FRM). Each module plays a critical role in enhancing the network's performance, and their synergistic interaction significantly improves the quality of super-resolution results for infrared images. We now provide a detailed description of these three modules.

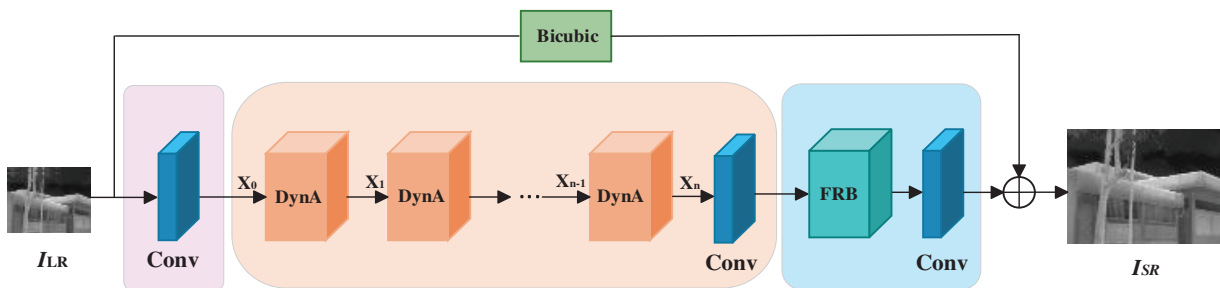


Figure 1: Structure of LDynSR

Given an LR Infrared image $I_{LR} \in R^{C \times H \times W}$, the input image is first passed through a 3×3 convolutional layer for shallow feature extraction. The module can be represented as:

$$x_0 = f_{conv}(I_{LR}) \quad (1)$$

where $f_{conv}(\cdot)$ denotes the 3×3 convolution function and x_0 is the output feature extracted by this function. For lightweight design, only one convolutional layer is utilized in the shallow feature extraction part. This approach effectively captures low-frequency information, providing a solid foundation for subsequent feature refinement while balancing network accuracy and computational efficiency.

While the shallow feature extraction efficiently captures low-frequency information, higher-level feature representations are essential for reconstructing fine details in the final output. To achieve this, the extracted shallow feature x_0 is input to a DFEB module, which consists of n Dynamic Weighting Module (DynA) modules and a 3×3 convolutional layer to extract a more powerful feature representation. We denote the proposed DynA as $f_{DA}(\cdot)$ and the deep feature extraction module can be represented as:

$$x_n = f_{conv}(f_{DA}^n(f_{DA}^{n-1}(\dots f_{DA}^0(x_0)\dots))) \quad (2)$$

where x_n is the output feature map of the n th DynA and $f_{conv}(\cdot)$ denotes the 3×3 convolution function. This module is the core of our network, consisting of multiple stacked DynA modules followed by a convolutional layer. Each DynA module dynamically adjusts the contributions of different attention branches based on the input features, thereby enhancing the model's ability to capture critical details and improve feature representation. The concluding convolutional layer integrates the refined features from the DynA modules, generating a robust and detailed feature representation.

After refining features through the DynA module, the extracted representations serve as input to the FRM module, which contains a feature reconstruction block (FRB) and a 3×3 convolutional layer. The FRM module comprises a transposed convolutional layer for up-sampling features to the target resolution and a final convolutional layer for refining the reconstructed image. The detailed structure of this module is described in [Section 3.3](#). In addition, we add a global connectivity path to double-triple interpolate the input image I_{LR} and finally obtain:

$$I_{SR} = F_{FRM}(x_n) + f_{up}(I_{LR}) \quad (3)$$

where $F_{FRM}(\cdot)$ denotes the feature reconstruction module, $f_{up}(\cdot)$ denotes the bicubic interpolation operation, and I_{SR} denotes the final output target SR image.

3.2 Dynamic Attention Module (DynA)

As shown in [Fig. 1](#), the output x_0 of the input image after a 3×3 convolutional layer is passed through a DFEB module consisting of n DynA to extract more advanced deep features. Each DynA comprises a dynamic weight module (DAM), two attention branches, and one non-attention branch, as seen in [Fig. 2](#).

3.2.1 Dynamic Weighting Module (DAM)

In traditional models, static attention mechanisms are commonly used, where attention is uniformly applied across all layers or blocks. This uniform application can lead to redundant computations and suboptimal performance, particularly when some attention modules do not contribute positively to the final output. Inspired by dynamic attention [35], we observe that not all attention modules enhance ultimate performance, and treating all attention blocks equally is not always the most advantageous approach. Therefore, we apply this learnable dynamic weighting module to our network architecture, designating it

as the Dynamic Weighting Module (DAM). In contrast to traditional networks, DAM dynamically adjusts attention weights based on input features, enabling the network to focus on beneficial features while suppressing less useful ones, thereby enhancing the network's greater potential. Specifically, the dynamic weighting module regulates the contributions from attention and non-attention branches by weighted summation. The module can be represented as:

$$x_{n+1} = f_{1 \times 1} (\omega_n^{PA} \times x_n^{PA} + \omega_n^{BTA} \times x_n^{BTA} + \omega_n^{NA} \times x_n^{NA}) + x_n \quad (4)$$

where x_n^{PA} is the output of the pixel attention branch, x_n^{BTA} is the output of the brightness-texture attention branch, and x_n^{NA} is the output of the non-attention branch. ω_n^{PA} , ω_n^{BTA} , and ω_n^{NA} are the corresponding weights of the three branches, which are computed by the network based on the input features, rather than artificially set as fixed values. $f_{1 \times 1}(\cdot)$ denotes a 1×1 convolution function.

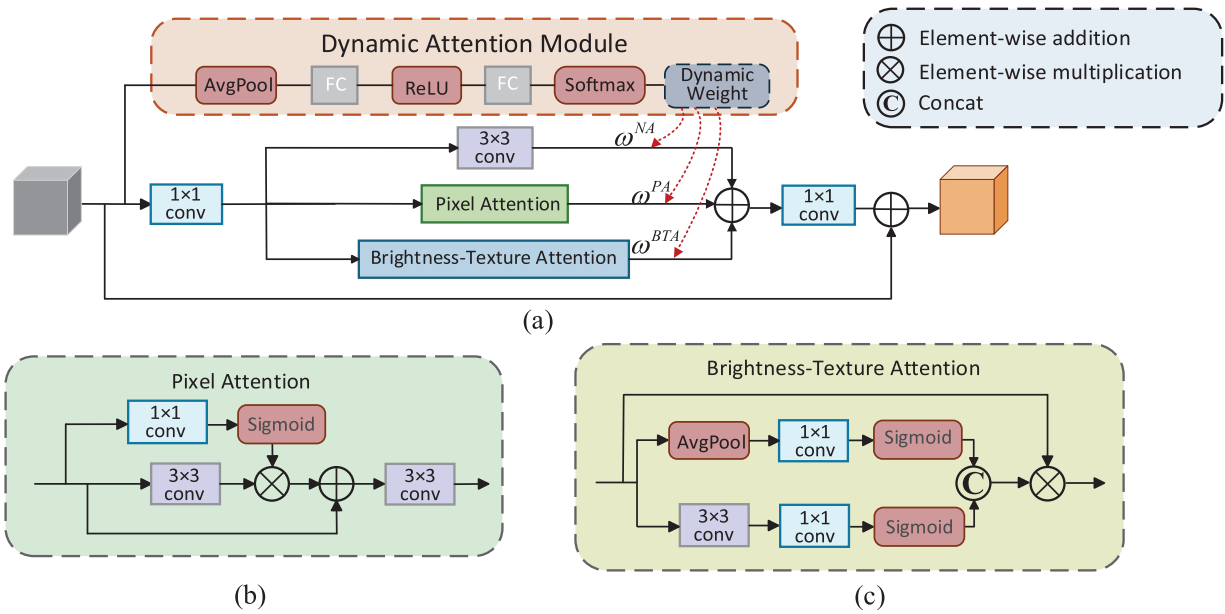


Figure 2: The overall structure of DynA and its attention mechanisms: (a) Structure of DynA. (b) Structure of pixel attention mechanism. (c) Structure of brightness-texture attention

To compute the dynamic weights, several key steps are involved, which are detailed as follows:

The input feature map x is first compressed using Global Average Pooling (GAP) to obtain global statistical information. This step aims to reduce the spatial dimensions of the feature map while preserving the channel-wise information. The process can be mathematically expressed as:

$$x_{pool} = \text{GlobalAveragePooling}(x) \quad (5)$$

where $x_{pool} \in R^{C \times 1 \times 1}$, and C is the number of channels.

After global average pooling, the features are passed through two fully connected layers equipped with ReLU activation functions to generate dynamic weights. Specifically, the first fully connected layer reduces the dimensionality of the global descriptor, projecting it into a lower-dimensional intermediate space. This reduction helps to lower computational complexity while retaining essential feature information. Subsequently, the second fully connected layer maps the intermediate representation back to the original

dimension, yielding the final weights that signify the importance of each branch. This process can be expressed as:

$$x_{fc1} = \text{ReLU}(W_1 \cdot x_{pool} + b_1) \quad (6)$$

$$x_{fc2} = W_2 \cdot x_{fc1} + b_2 \quad (7)$$

where W_1 and W_2 are the weight matrices of the fully connected layers, and b_1 and b_2 are the bias terms.

After passing through the fully connected layers, the output is normalized by the Softmax function, generating three weights ω_n^{PA} , ω_n^{BTA} , and ω_n^{NA} , which correspond to the weights of the PA attention branch, BTA attention branch, and non-attention branch, respectively. The formula is expressed as:

$$\omega_n^{PA}, \omega_n^{BTA}, \omega_n^{NA} = \text{Softmax}(x_{fc2}) \quad (8)$$

where $\omega_n^{PA} + \omega_n^{BTA} + \omega_n^{NA} = 1$, ensuring that the sum of the weights of the three branches is 1.

This module adaptively adjusts the branch weights according to the input features, supports resource allocation of multi-branch networks, and helps the model handle different inputs more flexibly. Once the input features change, the weight of each branch will also change, thereby improving the feature expression ability.

3.2.2 Attention Branches

[Fig. 2](#) illustrates that the attention branch comprises pixel attention (PA) and brightness-texture attention (BTA). Their structures are depicted in (b) and (c) of [Fig. 2](#), respectively.

To enhance the feature representation capability of the pixel attention mechanism, we propose a PA mechanism with dynamic feature enhancement. The proposed PA incorporates residual connections based on dynamically enhanced features, thereby achieving the enhancement of salient features and the effective preservation of original information. The module first generates pixel-level dynamic weights through a lightweight 1×1 convolution and adjusts the saliency of features extracted by a 3×3 convolution through point-by-point multiplication. In addition, the residual connection stabilizes the training process while preserving the global characteristics, avoiding the problem of gradient vanishing or information loss. It not only improves the model's capacity to capture multi-scale features but also substantially improves the robustness and performance of the model.

To enhance the super-resolution performance of infrared images while reducing computational complexity, we propose a Brightness-Texture Attention (BTA) mechanism. This module dynamically adjusts the significance of different regions in the image by introducing adaptive attention mechanisms for brightness and texture. Specifically, we design two branches to extract brightness information and texture information, respectively. For brightness information, we employ global average pooling to extract brightness features, while the other branch extracts texture features through depth-separable convolution, which significantly reduces the computational overhead. Meanwhile, the attention map combined with sigmoid activation enhances the recovery of image details, especially in infrared images with low contrast and texture loss, significantly enhancing the resolution efficacy of the image. Moreover, the lightweight design of the module enables it to operate in environments with limited computational resources and adapt to large-scale real-world applications, and it can achieve significant performance enhancement in the super-resolution task of infrared images while maintaining low computational complexity.

3.3 Feature Reconstruction Block (FRB)

In the feature reconstruction module, the core goal is to progressively enlarge the feature map to the target resolution to provide high-quality feature support for image reconstruction. To this end, we present a feature reconstruction module utilizing multi-stage upsampling and an attention mechanism, which applies varying processing strategies based on the scale factors to enhance the quality of super-resolution reconstruction, as shown in Fig. 3.

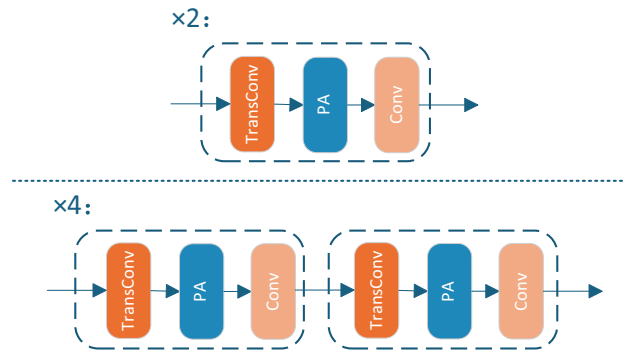


Figure 3: Structure of FRM

When the scale factor is 2, a single transposed convolutional layer (TransConv) is employed for upsampling instead of the conventional nearest-neighbor interpolation. Compared to traditional upsampling methods, transposed convolution transforms the upsampling process from a fixed interpolation pattern into a learnable procedure, enabling the model to adaptively refine texture and detail representations based on input image features. This learnable upsampling process facilitates more accurate feature recovery, particularly for complex and diverse structures. Subsequently, the up-sampled feature maps are refined by a PA layer to enhance the representation of important regions, followed by the application of a convolutional layer to restore the high-resolution details of the image. When the scale factor is 4, the super-resolution task is more burdensome. To mitigate the distortion problem associated with direct quadruple magnification, we adopt a two-stage up-sampling strategy. This approach progressively refines the features and generates a high-resolution output image, which continuously enhances the feature representation ability during the gradual magnification process and enables more effective recovery of image details.

4 Experiments

4.1 Datasets

We utilize the grayscale infrared dataset introduced by Rivadeneira et al. [36], which contains 951 training images and 50 test images. In the training process, we use 951 FLIR (HR) images as the original HR training dataset, use $\times 2$ and $\times 4$ bicubic downsampling to create the corresponding LR datasets, and degrade them by additive white Gaussian noise (AWGN) with a mean of 0 and a standard deviation of 10. To evaluate the generalization of our proposed network on the test dataset, we utilize not only the 50 test sets in FLIR (HR) in the challenge dataset but also use 101ThermalTau2 [23] (containing 101 grayscale infrared datasets of different scenes) for testing.

4.2 Evaluation Metrics

To quantitatively evaluate the performance of the proposed lightweight dynamic super-resolution method (LDynSR) for infrared images, we adopt widely used image quality assessment metrics: Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM). All performance reports are evaluated on the Y-channel (luminance) of the Luma, Chroma-blue, Chroma-red (YCbCr) color space, the reason being that the Y-channel mainly contains the brightness information of the image, which directly affects visual clarity and detail perception. PSNR is a classic metric for assessing image restoration quality, and its calculation is based on pixel-level error, with higher values signifying superior image restoration. Nevertheless, PSNR is not sensitive enough to the structural information of an image, so it is usually used in conjunction with SSIM, which can better reflect the visual quality of an image by evaluating its brightness, contrast, and structural information. Since infrared images usually lack rich color information, the preservation of structural information is particularly important, so we emphasize the performance of SSIM on the Y-channel.

4.3 Implementation Details

We use the PyTorch framework to train the proposed LDynSR. The Adam optimizer (where $\beta_1 = 0.9$, $\beta_2 = 0.99$, and $\epsilon = 10^{-8}$) and the L1 loss function are employed to train the model, while a cosine annealing learning approach is utilized to enhance training speed. The initial learning rate is 7×10^{-4} with 120 k iterations, and the cycle length in the cosine annealing learning rate schedule is 30 k iterations. The initial learning rate of 7×10^{-4} was empirically determined to balance convergence speed and stability. A higher rate risks overshooting the optimal solution, leading to suboptimal performance or divergence, while a lower rate slows training, increasing computational cost and time. The total number of iterations is 120 k, ensuring sufficient learning capacity without overfitting. The cycle length of 30 k iterations in the cosine annealing schedule was chosen to allow periodic restarts, which help escape local minima and improve generalization. We performed a sensitivity analysis by varying the cycle length and learning rate and observed that this configuration achieved the best trade-off between convergence speed and final performance. Data augmentation is applied to the training dataset by random rotations of 90° , 180° , and 270° , as well as horizontal flips. The proposed algorithm is implemented on a server equipped with an NVIDIA RTX 4090 GPU.

4.4 Ablation Experiments

Structural Configuration of Attention Branches: To assess the efficacy of the attention mechanism and the path configuration that we proposed in the DynA module on the super-resolution of infrared images, we set up seven sets of comparison experiments, which can be categorized into the single-path structure, two-path structure, and three-path structure. The single-path structure is NOA, PA, and BTA; the two-path structure contains PA-NOA, BTA-NOA, and PA-BTA, and the three-path structure is PA-BTA-NOA, where NOA denotes the non-attentive branch and consists of a 3×3 convolutional layer. The experimental results are shown in [Table 1](#).

As can be seen in [Table 1](#), in the single-path structure, using only the PA attention mechanism gives the best results, followed by using only our proposed BTA, which also has a low number of parameters of 71,883, proving that it is an attention structure with a lightweight structure, whereas the non-attention branch using only one 3×3 convolutional layer gives the worst results. In the dual-path structure, using the PA-BTA structure outperforms the PA-NOA structure, indicating that PA and BTA combine pixel-level features with brightness texture features, respectively. The amalgamation of these two can be used to achieve a more desirable effect and better performance, with the parameter count increasing by merely 1792 as compared to that of the PA-NOA. Meanwhile, the dual-path structure generally has a higher performance than the

single-path structure, which also indicates that in the dual-path structure, the feature information between each branch can complement each other. From Table 1, it can be seen that the three-path structure is the best, with a PSNR of 41.60, and the number of parameters does not increase dramatically, which achieves a balance between lightweight and high performance.

Table 1: Comparison of different attention branching structure configurations. Experiments were tested on the 101 ThermalTau2 dataset on a $\times 2$ scale factor

Case	Params	PSNR (dB)
Single-path structure		
NOA	125,931	41.303974
PA	190,251	41.334499
BTA	71,883	41.304022
Two-path structure		
PA-NOA	273,531	41.429606
BTA-NOA	155,163	41.396350
PA-BTA	275,323	41.487632
Three-path structure		
PA-BTA-NOA	277,163	41.604376

Effectiveness of PA in FRM module: We set up four sets of experiments in the feature reconstruction module using BTA, Channel Attention (CA), a 3×3 convolutional layer, and our proposed PA attention mechanism, respectively, for comparison tests to validate the effectiveness of our proposed method. Where CA is the channel attention module, the structure is shown in Fig. 4.

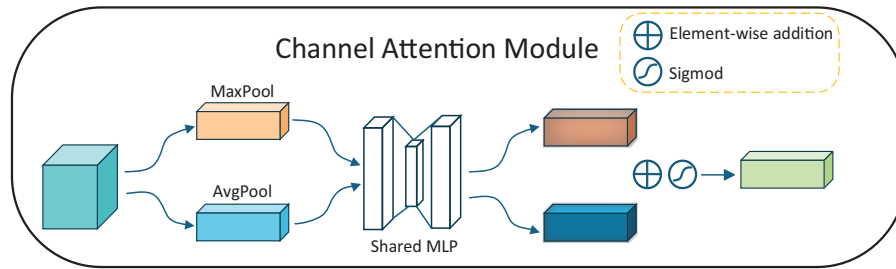


Figure 4: Structure of CA

The experimental results are presented in Table 2. The incorporation of the PA attention mechanism into the feature reconstruction module yields optimal results, and an improvement of 0.12 dB can be achieved compared to a 3×3 convolutional layer, which verifies that our proposed PA attention mechanism applied to the feature reconstruction module can bring beneficial effects for achieving super-resolution of infrared images.

Effectiveness of DAM: To assess the effectiveness of the DAM module, we conducted two sets of experiments: one group without DAM and another with DAM. The corresponding results are summarized in Table 3. The table indicates that the implementation of DAM can enhance model performance, compared

with not using the DAM module, bringing 0.08 dB improvement. It is also verified that the attention layer does not bring beneficial effects at different depths.

Table 2: Comparison of different attentional mechanism configurations in the FRM module. Experiments were tested on the 101 ThermalTau2 dataset on a $\times 2$ scale factor

Case	Params	PSNR (dB)
BTA	266,461	41.464360
CA	266,339	41.505732
3×3	271,403	41.583916
PA	277,163	41.604376

Table 3: Comparison of whether to use DAM or not. Experiments were tested on the 101 ThermalTau2 dataset on a $\times 2$ scale factor

Case	Params	PSNR (dB)
Without DAM	275,323	41.524879
With DAM	277,163	41.604376

4.5 Quantitative Experiments

The proposed method is evaluated against other existing state-of-the-art methods in terms of PSNR and SSIM at upscaling factors of $\times 2$ and $\times 4$. To ensure an even comparison, we incorporate many approaches, including Bicubic, SRCNN (2015) [9], Linearly-Assembled Pixel-Adaptive Regression Network (LAPAR_B) (2020) [37], FSRCNN (2016) [10], VDSR (2016) [11], PAN (2020) [20], IMDN (2019) [18], Channel-Shuffle Feature Mixing Network (ShuffleMixer) (2022) [38], and Lightweight Infrared Image Super-Resolution Network (LIRSRN) (2024) [39] with the test results presented in Table 4. Table 4 compares the performance and parameter count of various image super-resolution algorithms on the FLIR and Thermal101 datasets. The results indicate that our proposed LDynSR achieves the highest PSNR and SSIM at both $\times 2$ and $\times 4$ scale factors, and obtains SR images with the highest quantitative fidelity, demonstrating excellent performance and the superiority of the lightweight design. This outstanding performance can be attributed to several key design aspects of LDynSR, particularly the brightness-texture attention mechanism. Since infrared images primarily rely on brightness information and lack rich color information, effectively utilizing brightness features is crucial for infrared image super-resolution. The proposed brightness-texture attention mechanism adaptively adjusts the brightness weights, maintaining an appropriate contrast between high-brightness and low-brightness regions, thus enhancing the clarity of the target. Considering that the texture information in infrared images is often weak, leading to blurry or missing textures in traditional SR methods, the proposed mechanism enhances the texture details in key regions, resulting in clearer super-resolved images. In addition, due to our lightweight architecture—such as the use of a single 3×3 convolutional layer for shallow feature extraction, the lightweight attention mechanism, and the simplicity of the structure—our model uses fewer parameters compared to other methods. For example, compared with IMDN (0.68 M parameters for a scale factor of $\times 2$ and 0.70 M for a scale factor of $\times 4$) and VDSR (0.67 M), our proposed LDynSR (0.26 M parameters for a scale factor of $\times 2$ and 0.28 M for a scale factor of $\times 4$) achieves better performance using less than half of their parameters.

Table 4: Quantitative comparison results of the proposed method with other methods on PSNR (dB)/SSIM metrics

Scale	Network	Params	Testing datasets	
			FLIR	Thermal101
$\times 2$	Bicubic	–	35.83/0.9168	37.88/0.9057
	SRCNN	0.057 M	39.22/0.9570	40.67/0.9407
	LAPAR_B	0.24 M	39.37/0.9575	39.73/0.9252
	FSRCNN	0.012 M	40.15/0.9601	41.02/0.9438
	VDSR	0.67 M	40.26/0.9345	40.82/0.9422
	PAN	0.25 M	39.36/0.9649	40.54/0.9342
	IMDN	0.68 M	40.32/0.9651	41.47/0.9447
	ShuffleMixer	0.39 M	40.35/0.9596	41.45/0.9438
	LIRSRN	0.04 M	39.25/0.9436	39.47/0.9310
	LDynSR (ours)	0.26 M	40.44/0.9664	41.60/0.9618
$\times 4$	Bicubic	–	32.89/0.7788	35.21/0.8241
	SRCNN	0.057 M	35.06/0.9104	36.15/0.8957
	LAPAR_B	0.31 M	34.80/0.8953	36.77/0.8910
	FSRCNN	0.012 M	35.13/0.9025	36.52/0.9063
	VDSR	0.67 M	35.34/0.8997	36.30/0.8994
	PAN	0.25 M	35.18/0.8913	35.26/0.8832
	IMDN	0.70 M	35.42/0.9002	36.69/0.9018
	ShuffleMixer	0.40 M	35.45/0.9028	36.77/0.9015
	LIRSRN	0.04 M	34.64/0.8945	35.35/0.8803
	LDynSR (ours)	0.28 M	35.50/0.9033	36.86/0.9024

4.6 Qualitative Experiments

The qualitative results are presented in Figs. 5 and 6, where experiments were performed on the FLIR and Thermal101 test datasets with scale factors of $\times 2$ and $\times 4$, respectively. Figs. 5a and 6a present the selected test images, while Figs. 5b and 6b highlight the corresponding local details. The results of various super-resolution methods, including the proposed LDynSR, are shown in Figs. 5c–l and 6c–l. As observed, LDynSR exhibits superior performance in terms of PSNR and SSIM, and better preserves high-frequency details compared to other competing approaches. This improvement is particularly evident in fine-texture regions, where the LDynSR effectively recovers intricate details such as edges and small structures that are often lost in other super-resolution methods. For example, as shown in Fig. 6, at a scaling factor of $\times 4$, our method LDynSR (Fig. 6l) correctly restores the direction of the stripes in the image, while the FSRCNN (Fig. 6f) method causes distortion in the stripes during the restoration process. Compared with the PAN (Fig. 6h) method, our approach produces better contrast and is closer to the original HR image. This can be attributed to the synergistic effect of the pixel attention mechanism, which focuses on critical areas of the image, and the brightness-texture attention mechanism, which refines the extraction of both global and local image features. In addition, the non-attention mechanism supplements features that might otherwise be overlooked, contributing to the overall enhancement of image quality. These qualitative results validate the effectiveness of our network in achieving high-quality infrared image super-resolution.

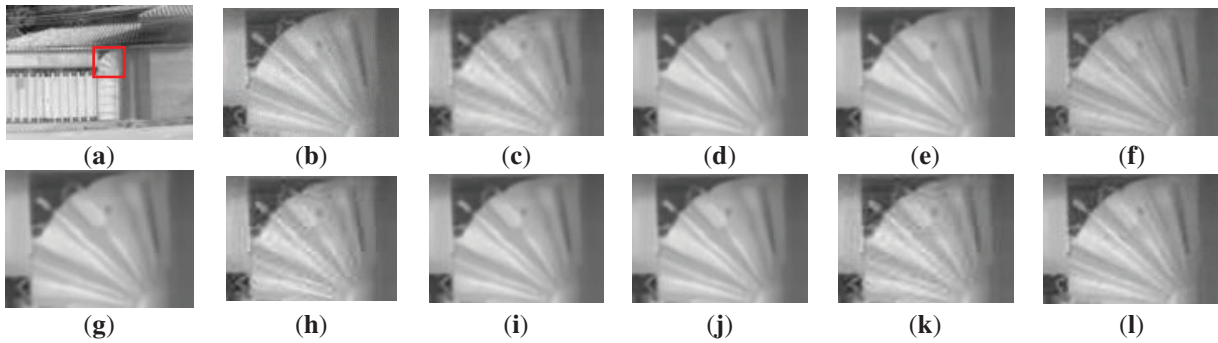


Figure 5: Qualitative results for the FLIR test dataset with a scale factor of $\times 2$: (a) 0046.jpg from FLIR; (b) HR PSNR (dB)/SSIM; (c) Bicubic 30.26/0.9430; (d) SRCNN 35.88/0.9804; (e) LAPAR_B 35.76/0.9798; (f) FSRCNN 35.53/0.9841; (g) VDSR 35.98/0.9809; (h) PAN 34.82/0.9669; (i) IMDN 36.54/0.9763; (j) ShuffleMixer 36.87/0.9874; (k) LIRSRN 34.49/0.9775; (l) LDynSR (ours) 37.41/0.9884

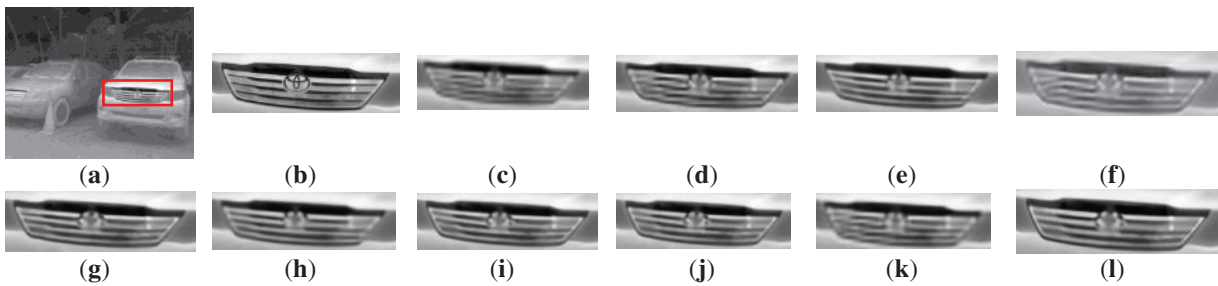


Figure 6: Qualitative results for the Thermal101 test dataset with a scale factor of $\times 4$: (a) thermal_055-S.png from Thermal101; (b) HR PSNR (dB)/SSIM; (c) Bicubic 28.52/0.9358; (d) SRCNN 33.44/0.9689; (e) LAPAR_B 33.80/0.9646; (f) FSRCNN 31.53/0.9541; (g) VDSR 34.96/0.9759; (h) PAN 33.99/0.9653; (i) IMDN 35.00/0.9758; (j) ShuffleMixer 35.22/0.9763; (k) LIRSRN 33.34/0.9426; (l) LDynSR (ours) 35.46/0.9771

5 Conclusions

In this paper, we propose LDynSR, a lightweight super-resolution network for infrared images that leverages an adaptive attention mechanism. To optimize the extraction of pixel-level information from infrared images, we developed a novel pixel attention mechanism (PA) incorporating dynamic feature augmentation. Additionally, we introduce residual connectivity on top of dynamic feature enhancement, which effectively enhances salient features while preserving original information. Furthermore, we propose a brightness-texture attention (BTA) mechanism to better capture the brightness-texture information of infrared images, thereby improving the recovery of image details, especially in infrared images with low contrast and texture loss, and significantly enhancing their super-resolution performance. Finally, we employ a dynamic weighting module to adjust the contributions of the attention layers, fully utilizing these attention branches and sub-attention branches to enhance high-frequency detail extraction, thus achieving superior performance. In the feature reconstruction module, we employ transposed convolution for upsampling to more accurately recover image details and reduce artifacts caused by interpolation. Qualitative and quantitative experiments conducted on two test datasets encompassing diverse scenarios demonstrate LDynSR's superior ability to reconstruct high-frequency details while preserving fine features. Furthermore, this lightweight architecture has the potential to be deployed on edge devices. Future research will focus on exploring further optimization of the adaptive weighting mechanism to improve the balance between attention and non-attention branches, potentially enhancing feature extraction efficiency. Additionally,

integrating knowledge distillation or pruning techniques could further reduce computational complexity, making the model more suitable for real-time applications on resource-constrained devices. Finally, the application of LDynSR to other types of images, such as medical imaging, could be explored to enhance its robustness and generalizability.

Acknowledgement: We would like to thank all individuals and organizations that provided support throughout this research. Their contributions have been invaluable to the completion of this work.

Funding Statement: This research was funded in part by the Henan Province Key R&D Program Project, “Research and Application Demonstration of Class II Superlattice Medium Wave High Temperature Infrared Detector Technology” under Grant No. 231111210400.

Author Contributions: Conceptualization and methodology, Yong Gan and Mengke Tang; software, Mengke Tang; validation, Xinxin Gan and Mengke Tang; formal analysis, Mengke Tang; resources and data curation, Yifan Zhang and Xinxin Gan; writing—original draft preparation, Mengke Tang and Yong Gan; writing—review and editing, Yong Gan, Mengke Tang, Xinxin Gan and Yifan Zhang; visualization, Xinxin Gan; supervision, Yong Gan; project administration, Yong Gan. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: The data that support the findings of this study are available from the corresponding author, Yong Gan, upon reasonable request.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest to report regarding the present study.

References

1. Chen H, He X, Qing L, Wu Y, Ren C, Sheriff RE, et al. Real-world single image super-resolution: a brief review. *Inf Fusion*. 2022;79:124–45. doi:10.1016/j.inffus.2021.09.005.
2. Laikin M. *Lens design*. Boca Raton, FL, USA: CRC Press; 2018.
3. Wang Z, Chen J, Hoi SC. Deep learning for image super-resolution: a survey. *IEEE Trans Pattern Anal Mach Intell*. 2020;43(10):3365–87. doi:10.1109/tpami.2020.2982166.
4. Liu TJ, Chen YZ. Satellite image super-resolution by 2D RRDB and edge-enhanced generative adversarial network. *Appl Sci*. 2022;12(23):12311. doi:10.3390/app122312311.
5. Zhang X, Huang W, Xu M, Jia S, Xu X, Li F, et al. Super-resolution imaging for infrared micro-scanning optical system. *Opt Express*. 2019;27(5):7719–37. doi:10.1364/oe.27.007719.
6. Huang JB, Singh A, Ahuja N. Single image super-resolution from transformed self-exemplars. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*; 2015 Jun 7–12; New York, NY, USA: IEEE; 2015.
7. Timofte R, De Smet V, Van Gool L. Anchored neighborhood regression for fast example-based super-resolution. In: *Proceedings of the IEEE International Conference on Computer Vision*; 2013 Dec 1–8; New York, NY, USA: IEEE; 2014.
8. Yang J, Wright J, Huang TS, Ma Y. Image super-resolution via sparse representation. *IEEE Trans Image Process*. 2010;19(11):2861–73. doi:10.1109/tip.2010.2050625.
9. Dong C, Loy CC, He K, Tang X. Image super-resolution using deep convolutional networks. *IEEE Trans Pattern Anal Mach Intell*. 2015;38(2):295–307. doi:10.1109/tpami.2015.2439281.
10. Dong C, Loy CC, Tang X. Accelerating the super-resolution convolutional neural network. In: *Proceedings of the Computer Vision–ECCV 2016: 14th European Conference*; 2016 Oct 11–14; Amsterdam, The Netherlands. Cham, Switzerland: Springer; 2016.

11. Kim J, Lee JK, Lee KM. Accurate image super-resolution using very deep convolutional networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2016 Jun 27–30; New York, NY, USA: IEEE; 2016.
12. Zou L, Xu S, Zhu W, Huang X, Lei Z, He K. Improved generative adversarial network for super-resolution reconstruction of coal photomicrographs. *Sensors*. 2023;23(16):7296. doi:10.3390/s23167296.
13. Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. *Adv Neural Inf Process Syst*. 2012;25(6):1097–105. doi:10.1145/3065386.
14. Lim B, Son S, Kim H, Nah S, Mu Lee K. Enhanced deep residual networks for single image super-resolution. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops; 2017 Jul 21–26; New York, NY, USA: IEEE; 2017.
15. Zhang Y, Li K, Li K, Wang L, Zhong B, Fu Y. Image super-resolution using very deep residual channel attention networks. In: Proceedings of the European Conference on Computer Vision (ECCV); 2018 Sep 08–14; Cham, Switzerland: Springer; 2018.
16. Kim J, Lee JK, Lee KM. Deeply-recursive convolutional network for image super-resolution. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2016 Jun 26–30; New York, NY, USA: IEEE; 2016.
17. Ahn N, Kang B, Sohn KA. Fast, accurate, and lightweight super-resolution with cascading residual network. In: Proceedings of the European Conference on Computer Vision (ECCV); 2018 Sep 8–14; Cham, Switzerland: Springer; 2018.
18. Hui Z, Gao X, Yang Y, Wang X. Lightweight image super-resolution with information multi-distillation network. In: Proceedings of the 27th ACM International Conference on Multimedia; 2019 Oct 21–25; New York, NY, USA: Association for Computing Machinery; 2019.
19. Liu J, Tang J, Wu G. Residual feature distillation network for lightweight image super-resolution. In: Proceedings of the Computer Vision–ECCV 2020 Workshops; 2020 Aug 23–28; Glasgow, UK. Cham, Switzerland: Springer; 2020.
20. Zhao H, Kong X, He J, Qiao Y, Dong C. Efficient image super-resolution using pixel attention. In: Proceedings of the Computer Vision–ECCV 2020 Workshops; 2020 Aug 23–28; Glasgow, UK. Cham, Switzerland: Springer; 2020.
21. Liu J-J, Hou Q, Cheng MM, Wang C, Feng J. Improving convolutional networks with self-calibrated convolutions. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition; 2020 Jun 13–19; New York, NY, USA: IEEE; 2020.
22. Choi Y, Kim N, Hwang S, Kweon IS. Thermal image enhancement using convolutional neural network. In: Proceedings of the 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS); 2016 Oct 09–10; New York, NY, USA: IEEE; 2016.
23. Rivadeneira RE, Suárez PL, Sappa AD, Vintimilla BX. Thermal image superresolution through deep convolutional neural network. In: Proceedings of the Image Analysis and Recognition: 16th International Conference (ICIAR 2019); 2019 Aug 27–29; Waterloo, ON, Canada. Cham, Switzerland: Springer; 2019.
24. Chudasama V, Patel H, Prajapati K, Upla KP, Ramachandra R, Raja K, et al. Therisurnet-a computationally efficient thermal image super-resolution network. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops; 2020 Jun 14–19; New York, NY, USA: IEEE; 2020.
25. Prajapati K, Chudasama V, Patel H, Sarvaiya A, Upla KP, Raja K, et al. Channel split convolutional neural network (ChaSNet) for thermal image super-resolution. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition; 2021 Jun 19–25; New York, NY, USA: IEEE; 2021.
26. Zou Y, Zhang L, Liu C, Wang B, Hu Y, Chen Q. Super-resolution reconstruction of infrared images based on a convolutional neural network with skip connections. *Opt Lasers Eng*. 2021;146(18):106717. doi:10.1016/j.optlaseng.2021.106717.
27. Du Y-B, Sun H-M, Zhang B, Cui Z, Jia R-S. A multi-scale mixed convolutional network for infrared image super-resolution reconstruction. *Multimed Tools Appl*. 2023;82(27):41895–911. doi:10.1007/s11042-023-15359-0.
28. Wu H, Hao X, Wu J, Xiao H, He C, Yin S. Deep learning-based image super-resolution restoration for mobile infrared imaging system. *Infrared Phys Technol*. 2023;132(6):104762. doi:10.1016/j.infrared.2023.104762.

29. Hu J, Shen L, Sun G. Squeeze-and-excitation networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2018 Jun 18–13; New York, NY, USA: IEEE; 2018.
30. Roy AG, Navab N, Wachinger C. Concurrent spatial and channel ‘squeeze & excitation’ in fully convolutional networks. In: Proceedings of the Medical Image Computing and Computer Assisted Intervention–MICCAI 2018: 21st International Conference; 2018 Sep 16–20; Granada, Spain. Berlin/Heidelberg, Germany: Springer; 2018.
31. Woo S, Park J, Lee J-Y, Kweon IS. CBAM: convolutional block attention module. In: Proceedings of the European Conference on Computer Vision (ECCV); 2018 Sep 8–14; Berlin/Heidelberg, Germany: Springer; 2018.
32. Fu J, Liu J, Tian H, Li Y, Bao Y, Fang Z, et al. Dual attention network for scene segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition; 2019 Jun 15–20; New York, NY, USA: IEEE; 2019.
33. Wang Q, Wu B, Zhu P, Li P, Zuo W, Hu Q. ECA-Net: efficient channel attention for deep convolutional neural networks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition; 2020 Jul 13–19; New York, NY, USA: IEEE; 2020.
34. Hu Y, Li J, Huang Y, Gao X. Channel-wise and spatial feature modulation network for single image super-resolution. *IEEE Trans Circuits Syst Video Technol.* 2019;30(11):3911–27. doi:10.1109/tcsvt.2019.2915238.
35. Chen H, Gu J, Zhang Z. Attention in attention network for image super-resolution. *arXiv:210409497*. 2021.
36. Rivadeneira RE, Sappa AD, Vintimilla BX. Thermal image super-resolution: a novel architecture and dataset. In: *VISIGRAPP (4: VISAPP)*; 2020 Feb 27–29; Cham, Switzerland: Springer; 2021.
37. Li W, Zhou K, Qi L, Jiang N, Lu J, Jia J. Lapar: linearly-assembled pixel-adaptive regression network for single image super-resolution and beyond. *Adv Neural Inf Process Syst.* 2020;33:20343–55.
38. Sun L, Pan J, Tang J. Shufflemixer: an efficient convnet for image super-resolution. *Adv Neural Inf Process Syst.* 2022;35:17314–26.
39. Lin C-A, Liu T-J, Liu K-H. LIRSRN: a lightweight infrared image super-resolution network. In: Proceedings of the 2024 IEEE International Symposium on Circuits and Systems (ISCAS); 2024 May 19–22; New York, NY, USA: IEEE; 2024.