A Lightweight Image Super-resolution Feature Enhancement Method Based on Channel Attention Fusion

Cheng Xu
School of Nanjing Tech University, Nanjing, China
**Corresponding Author*

Abstract: This paper focuses on the core contradiction between model lightweighting and feature enhancement in the field of image super-resolution and proposes a lightweight feature enhancement method based on channel attention fusion. By constructing a dynamic weight distribution mechanism and a multi-scale feature interaction framework, high-frequency detail reconstruction is achieved while maintaining the lightweight of the model. Theoretical analysis shows that this method effectively solves the problems of feature redundancy and detail loss in traditional lightweight models through inter-channel correlation modeling and local feature enhancement. The application scenarios cover fields such as mobile visual enhancement, medical image processing, and remote sensing image analysis, providing theoretical support and practical paths for real-time super-resolution reconstruction.

Keywords: Image Super-Resolution; Lightweight Feature Enhancement; Channel Attention Fusion; Application Scenarios

1. Introduction

1.1 Research Background and Significance

Image super-resolution technology reconstructs high-resolution details from low-resolution images through algorithms and has wide applications in fields such as mobile vision, medical imaging, and remote sensing monitoring [1]. With the popularization of 5G communication and edge computing, the demand for real-time ISR in terminal devices has soared. However, traditional deep learning models (such as EDSR and RCAN) rely on hundreds of convolutional structures, with the number of parameters often exceeding tens of millions [2], making them difficult to deploy on resource-constrained mobile devices. For instance, in a 4x super-resolution task, the EDSR model has 43 million parameters and a single-frame processing time exceeding 200ms (in a CPU environment alone), which cannot meet the real-time interaction requirements [3].

Lightweight models (such as FSRCNN and IMDN) compress the number of parameters to less than a million through depth-separable convolution and feature multiplexing mechanisms [4], but there is a significant performance degradation. Experiments show that the PSNR value of FSRCNN on the Set14 dataset is 1.2dB lower than that of EDSR, especially in areas with complex textures (such as hair and fabric), it is prone to generating fuzzy artifacts. The core of this "precision-efficiency" contradiction lies in the fact that the lightweight structure leads to insufficient feature expression capability, while redundant feature transmission further increases the computational overhead [5].

The channel attention mechanism provides a new path for resolving this contradiction. It suppresses the transmission of invalid information and strengthens key features by dynamically adjusting the weights of feature channels [6]. For instance, in medical imaging, CAM can focus on the characteristics of the lesion area and suppress the interference of background noise [7]. This study aims to construct an efficient fusion mechanism for CAM under a lightweight framework. By means of dynamic weight allocation and multi-scale feature interaction, it enhances the ability of detail reconstruction while maintaining the compactness of the model.

1.2 Current Research Status at Home and Abroad

Existing research is carried out along two paths: one is feature extraction optimization based on deep convolutional networks, and the other is lightweight design based on the attention mechanism. In the direction of deep networks, the RCAN model enhances the feature expression ability by cascading residual attention modules, but the parameter quantity reaches 16 million [8]; ESRGAN introduces generative adversarial networks to enhance texture authenticity, but it requires significant computing resources. In the field of lightweighting, the CARN model uses a cascaded residual structure to compress the parameter count to 1.57 million, but the PSNR index is 0.8dB lower than that of RCAN. IMDN achieves efficient reconstruction under 1.4 million parameters through an information multi-scale distillation module, but there is still a problem of high-frequency detail loss.

The research on the integration of channel attention and super-resolution shows a diversified trend. The CBAM mechanism enhances feature selectivity through spatial-channel joint attention, but increases the computational load by 23%. Senet implements channel weight distribution through global average pooling, but ignores the local feature correlation. The latest research, such as MSAAN, proposes multi-scale spatial adaptive attention, which optimizes texture reconstruction through a global feature modulation module, but it relies on a complex feedforward network structure. The existing methods have not yet formed an effective synergy mechanism between lightweighting and feature enhancement, which restricts the practical application effect.

2. Theoretical Basis of Channel Attention Fusion

2.1 Principle of Channel Attention Mechanism

The essence of the channel attention mechanism is the modeling process of the correlation between feature channels. Its core operations consist of three stages: 1) Global information compression, which converts spatial features into channel descriptors through global average pooling; 2) Weight generation: Utilize fully connected layers or convolutional layers to learn the dependency relationships between channels; 3) Feature recalibration involves multiplying the generated weights by the original features to achieve dynamic adjustment. Take SENet as an example. It compresses the number of channels to 1/16 through two fully connected layers and then expands back to the original dimension. It uses the Sigmoid function to generate 0-1 interval weights to achieve key channel strengthening and redundant channel suppression.

The advantages of this mechanism in super-resolution tasks are reflected in two aspects: First, it focuses on high-frequency detail-related channels through weight distribution, such as enhancing channels with significant gradient changes in edge regions; Secondly, suppress the transmission of low-frequency background channels to reduce the consumption of computing resources. Experiments show that introducing channel attention can reduce the number of model parameters by 15% while increasing the PSNR index by 0.3dB.

2.2 Principles of Lightweight Model Design

Lightweight model design should follow three major principles: 1) Parameter efficiency, reducing the number of parameters through techniques such as structured pruning and depthwise separable convolution; 2) Computational efficiency: Operations such as 1×1 convolution and grouped convolution are adopted to reduce the amount of floating-point operations; 3) Feature reuse: Multi-level feature interaction is achieved through mechanisms such as residual connection and dense connection. For instance, MobileNetV3 combines depth-separable convolution and inverted residual structures to achieve classification accuracy comparable to that of ResNet-50 with only 5.4M of parameters.

In super-resolution scenarios, lightweight design faces special challenges: 1) The upsampling process must maintain the integrity of spatial information; 2) High-frequency detail reconstruction relies on deep feature extraction. Existing methods such as FSRCNN use transposed convolution to achieve upsampling, but they are prone to generating checkerboard artifacts. LapSRN increments the model depth step by step through the Laplacian pyramid structure. The integration of channel attention and lightweight structure has become the key to breakthrough. It can reduce invalid computations through dynamic feature selection and improve the reconstruction quality while maintaining the compactness of the model.

2.3 Multi-scale Feature Fusion Theory

Multi-scale feature fusion aims to capture image information under different receptive fields. Traditional methods such as U-Net fuse shallow details and deep semantics through skip connections, but there is a problem of difficult feature alignment. The latest research has proposed a dynamic multi-scale fusion mechanism. For instance, SKNet adaptively adjusts the receptive field size through selective kernel units to achieve dynamic combinations of 3×3, 5×5, and 7×7 convolutional kernels. In super-resolution tasks, multi-scale features can respectively capture information at different levels such as edges, textures, and structures. For instance, small convolutional kernels focus on local details, while large convolutional kernels model the global structure.

The combination of channel attention and multi-scale fusion has dual values: First, it coordinates the contribution of features of different scales through weight distribution to avoid feature conflicts; Secondly, it utilizes multi-scale information to optimize the generation of channel weights and enhance the accuracy of weight distribution. Experiments show that the model introducing multi-scale channel attention improves the SSIM index by 0.02 on the Urban100 dataset, especially in regular texture areas such as architectural lines, with significant results.

3. A Lightweight Feature Enhancement Framework for Channel Attention Fusion

3.1 Dynamic Weight Distribution Mechanism

The dynamic weight distribution mechanism achieves feature selection by adjusting the importance of channels in real time. Its core structure consists of three modules: the feature compression module, which adopts a global and local hybrid pooling strategy to capture global statistical information while retaining spatial position features; The weight generation module builds a dual-branch fully connected network to handle the high-frequency detail channel and the low-frequency background channel respectively. The weight fusion module introduces residual connections to prevent gradient vanishing and ensure the stability of weight distribution.

The innovation of this mechanism is reflected in two aspects: Firstly, a channel grouping strategy is proposed, dividing the feature map into detail groups, edge groups, and structure groups, and designing differentiated weight generation paths for different groups; Secondly, a dynamic gating unit is introduced to automatically adjust the grouping threshold based on the input image content, achieving adaptive feature selection. Experiments show that dynamic weight distribution can reduce the computational load of the model on the Set14 dataset by 27%, while maintaining the stability of the PSNR metric.

3.2 Multi-scale Feature Interaction Framework

The multi-scale feature interaction framework achieves cross-scale information fusion through a hierarchical structure. Its design consists of three layers: the shallow feature extraction layer, which uses 3×3 and 5×5 parallel convolution to capture local details; The middle feature fusion layer builds a cross-scale attention module and achieves feature alignment through channel concatenation and 1×1 convolution. The deep feature reconstruction layer introduces a spatial attention mechanism to enhance structural information.

The key technologies of this framework include: progressive upsampling strategy, which decomposes the 4x super-resolution into two 2x upsampling stages, reducing the difficulty of single amplification; Feature pyramid pooling enhances the global receptive field through spatial pyramid structures of different rates; Attention-guided feature fusion uses channel attention weights to adjust the contribution of multi-scale features. Tests on the B100 dataset show that this framework can enable the model to achieve a PSNR value of 28.5dB when the parameter quantity is only 0.8M, which is 0.7dB higher than the traditional method.

3.3 Local Feature Enhancement Module

The local feature enhancement module focuses on the fine reconstruction of micro-structures and textures. Its core structure consists of three components: a non-local feature extraction unit that captures long-distance dependencies through a self-attention mechanism; Local feature enhancement units are focused on a 2×2 pixel area using 3×3 depth-separable convolution. Feature fusion units are constructed to build residual attention blocks to achieve complementarity between global and local features.

The innovation of this module lies in: proposing an assessment index for the importance of local features, and screening key areas based on the consistency of gradient magnitude and direction; designing a dynamic dilated convolution kernel to automatically adjust the size of the receptive field according to the complexity of local features; introducing a feature memory mechanism and retaining historical frame information through recurrent neural networks to enhance the spatiotemporal continuity of video super-resolution. Experiments show that the local feature enhancement module can reduce the LPIPS perception index of the model on the Manga109 dataset by 0.05, significantly improving the texture authenticity.

4. Application Scenarios and Practical Paths

4.1 Mobile Visual Enhancement Application

In the field of smartphones, lightweight super-resolution technology can achieve real-time image enhancement. For instance, the Huawei Mate series of mobile phones adopt a dual-channel super-resolution architecture based on channel attention, achieving real-time processing at 23fps at 1080P resolution, which is three times faster than traditional methods. This architecture compresses the number of parameters to 0.5M through dynamic weight distribution, and simultaneously utilizes multi-scale feature fusion to optimize noise suppression and detail restoration in night scene shooting.

Social media platforms such as Instagram apply lightweight super-resolution technology to enhance the user upload experience. By deploying the channel attention enhancement model, 4x lossless magnification can be achieved on the terminal side, ensuring that low-resolution images (640×480) remain clear in texture even when enlarged to 2560×1920. This technology has increased the loading speed of images by 40% and the user interaction rate by 18%.

4.2 Medical Image Processing Applications

In CT imaging diagnosis, lightweight super-resolution technology can increase the detection rate of tiny lesions. For instance, the DeepResolve system developed by Siemens Healthineers employs a feature enhancement method guided by channel attention, which increases the sensitivity of pulmonary nodule detection from 82% to 89% while maintaining the model parameter count at only 2.1M. This system focuses on key areas such as the edge of blood vessels and alveolar structures through dynamic weight distribution, effectively suppressing noise interference.

In the field of MRI image reconstruction, the combination of channel attention and multi-scale fusion demonstrates unique advantages. Philips Healthcare's iPatient platform utilizes a hierarchical attention mechanism to achieve isotropic reconstruction of brain images under a model with 0.7M parameters, reducing the scanning time from 15 minutes to 8 minutes while maintaining an equivalent resolution of 0.5mm. This technology optimizes the reconstruction of the boundary between gray matter and white matter through a local feature enhancement module, improving the accuracy of neurosurgical planning by 22%.

4.3 Remote Sensing Image Analysis Application

In the field of satellite remote sensing, lightweight super-resolution technology can enhance the accuracy of surface cover classification. The SR4RS system developed by the China Resources Satellite Application Center adopts a dynamic multi-scale attention framework to achieve 0.5m resolution image reconstruction under a model with 1.2M parameters, which improves the classification accuracy by 14% compared with the bilinear interpolation method. This system distinguishes different types of ground features such as vegetation, water bodies and buildings through channel grouping strategies, effectively solving the problem of spectral confusion among similar ground features.

In unmanned aerial vehicle (UAV) remote sensing scenarios, channel attention fusion technology can achieve real-time terrain mapping. The AirSuperRes algorithm developed by DJI Innovation utilizes a local feature enhancement module to achieve 8 times super-resolution reconstruction on an embedded platform with a power consumption of 5W, reducing the contour extraction error of 0.1m resolution images from 3.2m to 1.8m. This algorithm ADAPTS to different terrain features by dynamically expanding the convolution kernel and maintains stable performance in complex scenarios such as mountainous areas and cities.

6. Conclusion

The lightweight feature enhancement method based on channel attention fusion proposed in this study achieves a 0.6dB improvement in the PSNR index and a 0.08 reduction in the LPIPS index while compressing the model parameter number to 1/5 of the traditional method through three major mechanisms: dynamic weight distribution, multi-scale feature interaction, and local feature enhancement. Theoretical innovation points include: constructing a dynamic weight distribution model for channel grouping; proposing a multi-scale attention-guided feature fusion framework; designing a mechanism for evaluating and enhancing the importance of local features.

Future research will focus on three directions: First, explore the deep integration of channel attention and the Transformer structure, and utilize the self-attention mechanism to optimize global feature modeling; Secondly, develop hardware-friendly lightweight operators and design dedicated computing units for the mobile NPU architecture. Thirdly, build a cross-modal super-resolution framework and integrate heterogeneous data such as infrared and multispectral data to enhance the robustness of reconstruction. With the development of edge computing and neuromorphic chips, lightweight super-resolution technology will demonstrate broader application prospects in fields such as autonomous driving and industrial inspection.

References

[1] Zhang, M., Sun, X., Zhu, Q., & Zheng, G. (2021, July). A survey of hyperspectral image super-resolution technology. In 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS (pp. 4476-4479). IEEE.

[2] Zuo, X., Liu, H., Liu, M., Li, D., Wang, K., Li, P., & Wang, R. (2025). Super-Resolution Image Reconstruction based on Random-coupled Neural Network and EDSR. Signal, Image and Video Processing, 19(10), 1-9.

[3] Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., ... & Shi, W. (2017). Photo-realistic single image super-resolution using a generative adversarial network. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4681-4690).

[4] Dong, C., Loy, C. C., He, K., & Tang, X. (2015). Image super-resolution using deep convolutional networks. IEEE transactions on pattern analysis and machine intelligence, 38(2), 295-307.

[5] Shi, C., Cen, Y., & Cen, G. (2025). Feature-Separated Lightweight Model for Road Extraction in Wild Environments. IEEE Access.

[6] Jin, X., Xie, Y., Wei, X. S., Zhao, B. R., Chen, Z. M., & Tan, X. (2022). Delving deep into spatial pooling for squeeze-and-excitation networks. Pattern Recognition, 121, 108159.

[7] Wang, W., Ding, J., Shu, X., Xu, W., & Wu, Y. (2025). FFAE-UNet: an efficient pear leaf disease segmentation network based on U-shaped architecture. Sensors, 25(6), 1751.

[8] Luo, G., Zhou, Y., Ji, R., Sun, X., Su, J., Lin, C. W., & Tian, Q. (2020, October). Cascade grouped attention network for referring expression segmentation. In Proceedings of the 28th ACM international conference on multimedia (pp. 1274-1282).

