Research on Image Super-Resolution Reconstruction Method Based on Deep Convolutional Neural Network

Ning Chen*

College of Artificial Intelligence, Guangzhou City Polytechnic, Guangzhou, 511370, China.
**Corresponding author: ningchen2024@126.com*

Abstract: *As an ill-posed inverse problem, image super-resolution reconstruction relies on effective modeling of natural image priors for its solution. Deep convolutional neural networks provide a new path to break through the limitations of traditional hand-crafted features. This paper systematically studies the theoretical framework of this direction: at the mathematical foundation level, it analyzes the degradation linear model and the ill-posedness of the inverse problem, and clarifies the transformation characteristics of convolution mapping and the constraints of receptive field; at the feature reuse level, it explores the gradient preservation of residual paths, the feature reuse of dense connections, and the recalibration mechanism driven by channel attention; at the reconstruction strategy level, it elaborates the Laplacian pyramid decomposition, the sub-pixel convolution shuffling operation, and the recursive back-projection error correction method. The above content constructs a theoretical framework of deep convolutional neural networks for super-resolution from three dimensions: network architecture, information flow, and multi-scale reconstruction.*

Keywords: *image super-resolution reconstruction; deep convolutional neural network; residual connection; dense connection; channel attention; sub-pixel convolution*

Introduction

Image super-resolution reconstruction is widely required in fields such as remote sensing, medicine, and surveillance. Its essence is an ill-posed inverse problem, which requires prior constraints to limit the solution space. Traditional interpolation and sparse coding are limited by hand-crafted features and tend to produce edge blurring and ringing artifacts. Deep convolutional neural networks learn the statistical laws of natural images in an end-to-end manner and integrate feature extraction and reconstruction mapping into a unified framework. The necessity of studying this method is reflected in the following aspects: the depth and width of the network affect the receptive field and the level of feature abstraction, thus requiring a reasonable design of connection modes to alleviate gradient attenuation; the upsampling strategy concerns the balance between accuracy and efficiency, and a single feedforward structure cannot handle multi-scale information. Systematically optimizing the behavior of deep convolutional neural networks from the three perspectives of mathematical constraints, feature reuse, and multi-scale reconstruction is of great significance for improving reconstruction quality and reducing model redundancy.

1. Mathematical Foundation and Model Constraints of Deep Convolutional Neural Network for Super-Resolution Reconstruction

1.1 Linear System Modeling of the Image Degradation Process and Ill-Posedness Analysis of the Inverse Problem

The image degradation process is typically described as a linear forward model in which a high-resolution image generates a low-resolution observation image after being subjected to blur kernel convolution, a downsampling operator, and additive noise interference. This model can be expressed as the low-resolution image being equal to the product of the degradation matrix and the high-resolution image vector plus the noise term, where the degradation matrix combines the joint effects of blurring and downsampling. Since the number of pixels in the high-resolution image is much larger than that in the low-resolution image, the degradation matrix exhibits a severe column-underdetermined property,

and its null space is nontrivial, which results in no unique solution for the inverse mapping. This ill-posedness manifests in three aspects: the existence of the solution is affected by noise perturbation; the uniqueness of the solution cannot be guaranteed due to information loss; and the stability of the solution is extremely sensitive to observation errors because of the rapid decay of the singular values of the degradation matrix.

To address the ill-posedness of the inverse problem, regularization methods introduce prior constraints into the objective function to shrink the solution space. From the perspective of functional analysis, a deep convolutional neural network can be viewed as an implicit regularizer: the network architecture and the parameterization process embed the high-resolution image manifold into a low-dimensional feature space, and the local connections and weight sharing of convolutional layers naturally limit the capacity of the hypothesis space. The nonlinear activation function further introduces sparsity constraints, enabling the network to learn a reconstruction mapping that conforms to the statistical characteristics of natural images. This implicit regularization does not require the explicit design of prior terms but automatically adapts to the degradation model through an end-to-end training process, thereby effectively mitigating the ill-posed nature of the inverse problem[1].

1.2 Transformation Characteristics of Convolution Operation in Mapping from Pixel Space to Feature Space

The convolution operation performs a weighted summation of pixels within a local neighborhood through a sliding window manner, thereby mapping the discrete lattice in the original pixel space to a multi-dimensional tensor representation in the feature space. This mapping possesses translation equivariance: a translation of the input image results in the same amount of translation in the output feature maps, and this property originates from the weight-sharing mechanism of the convolution kernel. From the perspective of the frequency domain, the Fourier transform of the convolution kernel determines its filtering characteristics. Convolution kernels of different sizes and parameters can respectively implement low-pass filtering to preserve structural contours or high-pass filtering to extract edges and texture details. The cascade of multiple convolutional layers is equivalent to a stepwise decomposition of the signal in the feature space, where the feature maps output by each layer correspond to the response components of the original image at different scales and orientations.

As the network depth increases, the feature space gradually deviates from the geometric structure of the original pixel domain and forms a hierarchical abstract representation. The shallow feature maps preserve the local correlations among pixels and mainly encode edges, corners, and simple texture information, and their transformation characteristics are close to those of a steerable filter bank. The deep feature maps combine local patterns into semantic concepts through multiple nonlinear transformations; as the spatial resolution decreases, the channel dimension increases, and the representation capability transitions from low-order statistics to high-order structural descriptions. This transformation process is not lossless, and some high-frequency details may be suppressed during the layer-by-layer mapping. The cross-layer connection mechanism delivers shallow features directly to deep layers through skip paths, which can compensate for the detail components lost during the feature space shift and maintain the multi-scale information integrity required for the reconstruction task[2].

1.3 Local Receptive Field Constraints and Global Context Dependency Trade-offs in Super-Resolution Reconstruction

In a standard convolutional neural network, the receptive field of each output neuron is limited by the product of the convolution kernel size and the number of network layers. For the super-resolution reconstruction task, the reconstruction of a single pixel depends not only on the local texture within its neighborhood but also on the self-similarity patterns of distant regions in the image. The local receptive field constraint makes it difficult for the model to capture large-scale repetitive structures, such as periodic textures or the continuity of long edges, thereby producing artifacts or discontinuities in the reconstruction results. When the receptive field is smaller than the support range of the degradation blur kernel, the model cannot perform correct deconvolution, which further exacerbates the reconstruction uncertainty.

To alleviate the contradiction between local constraints and global dependencies, dilated convolution can be adopted to exponentially expand the receptive field without increasing the number of parameters. The dilated convolution inserts holes between the convolution kernel elements, enabling a kernel of the same size to cover a larger input region while keeping the output feature map resolution

unchanged. An increasing combination of dilation rates can form a parallel structure of multi-scale receptive fields, allowing the network to perceive both local details and the global layout simultaneously. The introduction of the self-attention mechanism provides another approach: it computes the correlation weights between any pair of positions on the feature map and explicitly models non-local dependencies. This type of method can effectively suppress repetitive artifacts and discontinuous boundaries in the reconstructed image. However, it should be noted that dilated convolution may produce grid artifacts, and self-attention brings quadratic computational complexity. Through group dilation or sparse attention strategies, a reasonable trade-off between local and global constraints can be achieved without significantly sacrificing efficiency.

2. Feature Reuse Mechanisms in Deep Networks Based on Residual and Dense Connections

2.1 Gradient Preservation in Residual Paths and Convergence Enhancement of Identity Mapping

Deep convolutional networks are prone to the vanishing or exploding gradient phenomenon during the backpropagation process. The introduction of residual paths changes the information transmission mode of traditional sequential connections. The residual block passes the input directly through an identity mapping across the nonlinear transformation layers and then adds it to the output, which allows the gradient to be transmitted directly from deep layers to shallow layers without passing through the nonlinear activations and convolution operations of each layer. This shortcut connection constructs an unobstructed gradient highway, thereby alleviating the optimization difficulties caused by increasing depth. From the perspective of optimization theory, the residual structure transforms the mapping to be learned from the original direct fitting into a perturbation fitting of the identity mapping, that is, learning the difference between the input and the output. When the optimal mapping is close to the identity mapping, the residual branch only needs to learn a zero mapping, which is far less difficult than learning the identity mapping from scratch[3].

The preservation of the identity mapping relies on the element-wise addition operation between the skip connection and the main path output, which requires that the feature maps of the two branches have the same spatial size and number of channels. When the feature dimension changes, linear projection or zero-padding can be used to adjust the identity path; however, linear projection introduces additional parameters and may disrupt the integrity of direct gradient propagation. In deeper residual networks, the arrangement order of batch normalization and activation functions affects the purity of the identity mapping. The pre-activation residual unit places batch normalization and the activation function before the convolution layer, so that the identity path is free from any nonlinear transformation interference, further enhancing the smoothness of the gradient flow. This design ensures that the network maintains stable convergence behavior even when the depth increases to hundreds of layers, thus providing optimization feasibility for the large-capacity models required for super-resolution reconstruction.

2.2 Cross-Layer Feature Reuse and Parameter Efficiency Optimization in Dense Connection Structures

The dense connection network takes the output of each layer as the input to all subsequent layers, and each layer receives the set of feature maps from all previous layers and performs a concatenation operation along the channel dimension. This connection mode enables the network to achieve explicit cross-layer reuse in the feature dimension, so that the edge and texture information extracted by shallow layers can be directly delivered to deep layers without relying on the layer-by-layer transmission attenuation in residual learning. From the perspective of parameter efficiency, the dense connection controls the number of new feature maps added per layer through a narrow growth rate, so that the total parameter scale is much smaller than that of an ordinary convolutional network of the same depth. Each layer only needs to learn a small number of new features, and the remaining features reuse the computation results of previous layers, thus avoiding the redundant computation of repeatedly extracting the same feature across multiple layers.

In the super-resolution reconstruction task, the dense connection structure can effectively aggregate multi-level features ranging from local details to global structures. The shallow feature maps contain high-frequency edge information, the middle-level feature maps encode texture patterns, and the deep feature maps express semantic layout. The dense connection directly fuses these complementary features at the reconstruction stage. However, the dense connection also faces the storage overhead

problem that the number of feature map channels grows linearly with the number of layers. The compression operation in the transition layer reduces the feature dimension through convolution and pooling, thereby controlling memory usage while maintaining reuse efficiency. The connection mode between dense blocks also affects the effectiveness of feature reuse. Adopting a bottleneck structure can further compress the number of channels within each dense block, reducing the computational burden without losing the integrity of cross-layer information transmission.

2.3 Adaptive Feature Recalibration and Frequency Decomposition Driven by Channel Attention

Different feature channels undertake differentiated functional roles in the super-resolution reconstruction process: some channels are responsible for transmitting low-frequency structure information, while other channels focus on restoring high-frequency details. The channel attention mechanism compresses the spatial features of each channel into a scalar descriptor through global average pooling, and this descriptor reflects the global response strength of that channel. Subsequently, the mechanism generates a channel weight vector through a two-layer nonlinear transformation in a gating mechanism, and this vector performs channel-wise weighting on the original feature maps along the channel dimension. This process achieves adaptive recalibration of features, enhancing the channels that contribute more to the reconstruction task and suppressing the activation of redundant or irrelevant channels[4].

Combining channel attention with frequency decomposition can further optimize the band processing strategy in reconstruction. The low-frequency components correspond to the overall brightness and structural contours of the image, presenting a smooth response distribution in the feature maps, and their channel weights are usually high and stable; the high-frequency components correspond to edges and texture details, with sparse and localized responses that require finer weight modulation. The channel attention mechanism can implicitly learn this frequency band discrimination and control the retention degree of high-frequency information by adjusting the activation thresholds of different channels. Introducing a multi-branch frequency decomposition module into a deep network can decompose the input features into low-frequency approximation components and high-frequency residual components, and then send them into the attention module for independent processing. The low-frequency branch adopts a larger receptive field to maintain structural consistency, while the high-frequency branch adopts denser attention weights to enhance detail recovery capability. This frequency-aware attention strategy enables the network to adaptively adjust the feature response according to the frequency domain distribution of the input image, thereby improving the consistency of the reconstruction results in both objective and subjective metrics.

3. Multi-Scale Feature Extraction and Progressive Upsampling Reconstruction Strategy

3.1 Hierarchical Feature Decomposition and Fusion Guided by the Laplacian Pyramid

The Laplacian pyramid decomposes an image into residual components at different spatial frequency bands by successively downsampling and then upsampling and computing the difference. Each layer of the pyramid corresponds to a specific scale level, where the top layer contains the low-frequency approximation component of the image, and the lower layers contain high-frequency detail residuals. In super-resolution reconstruction, the hierarchical feature decomposition guided by the Laplacian pyramid splits the reconstruction task into multiple sub-problems: the low-frequency component performs structure recovery with a larger receptive field, while the high-frequency components perform detail prediction through shallower sub-networks. This decomposition strategy avoids the spectral aliasing phenomenon that occurs when a single network processes multi-band information simultaneously[5].

Hierarchical feature fusion occurs between different pyramid levels, adopting a top-down progressive information transmission path. The high-level low-frequency features are upsampled and then concatenated or element-wise added with the residual features of the next level to form a fused feature map. This fusion process is repeated until reaching the bottom level of the pyramid at the original image resolution. The multi-scale characteristics of the Laplacian pyramid naturally match the hierarchical feature extraction of convolutional networks: the shallow feature maps correspond to high-frequency residual components, and the deep feature maps correspond to low-frequency structure components. By designing cross-layer connections to map the internal features of the network to each level of the pyramid, the method can achieve end-to-end joint optimization, enabling the network to

automatically learn the optimal residual representation at different scales and improving the scale consistency of the reconstruction results.

3.2 Periodic Shuffling Operation in the Sub-Pixel Convolutional Layer and Spatial Resolution Enhancement

The traditional transposed convolution suffers from the checkerboard artifact problem during upsampling, and the root cause of this problem lies in the unevenness of the convolution kernel overlap pattern. The sub-pixel convolutional layer avoids this defect through a periodic shuffling operation: this layer performs ordinary convolution on the low-resolution feature map and outputs an intermediate feature map whose number of channels is the square of the upscaling factor, and then it rearranges the information from the channel dimension into the spatial dimension through pixel shuffling. The core of periodic shuffling is to concentrate the convolution computation in the low-resolution space and then increase the resolution through a fixed rearrangement mapping, so that the upsampling process does not introduce additional learnable parameter redundancy[6].

From the perspective of information theory, the sub-pixel convolutional layer treats upsampling as a mapping function from low-resolution features to high-resolution pixels, and this function has periodic local dependency characteristics. The value of each high-resolution pixel is jointly determined by multiple channels of information within its corresponding low-resolution neighborhood, and these channels encode information from different sub-pixel positions before shuffling. The periodic shuffling operation maintains the spatial continuity of the feature map and avoids the periodic artifacts caused by zero-padding in transposed convolution. In a super-resolution reconstruction network, the sub-pixel convolutional layer is usually placed at the end of the network and receives the multi-channel feature maps from the deep feature extraction module. By controlling the relationship between the number of channels of the intermediate feature map and the upscaling factor, the method can flexibly achieve an integer or non-integer multiple increase in spatial resolution while maintaining a balance between the visual quality of the reconstruction results and the computational efficiency.

3.3 Error Iterative Correction and Reconstruction Consistency Constraint in the Recursive Back-Projection Architecture

The recursive back-projection architecture draws on the error feedback principle in the iterative back-projection algorithm, and it achieves the gradual elimination of reconstruction errors by alternately performing an up-projection operation and a down-projection operation. The up-projection degrades the current estimated high-resolution image into a low-resolution residual, and the down-projection then back-projects this residual into the high-resolution space to update the estimation. Each back-projection operation generates a correction term, and this correction term is added to the current estimation to form a new high-resolution image. The recursive structure unrolls this iterative process into multiple network modules that share weights, and each module performs one complete back-projection update, with the modules connected sequentially to form a recursive chain.

The reconstruction consistency constraint requires that the final output high-resolution image, after undergoing the same degradation process, should remain consistent with the input low-resolution image. The recursive back-projection architecture calculates the difference between the degraded low-resolution image and the original input at each iteration, and it uses this difference as an error signal to guide the next correction. This internal consistency check mechanism ensures that the reconstruction result strictly conforms to the degradation model, thereby avoiding the texture drift or structural deformation that may occur in traditional feedforward networks. The recursion depth determines the number of error corrections; a deeper recursive structure can converge to a more accurate solution, but it should be noted that the gradient may face the problem of attenuation or explosion in the recursive chain. Adopting residual connections to directly add the input of each back-projection module to its output can stabilize the recursive training process while maintaining the progressive approximation characteristic of error correction.

Conclusion

This paper systematically studies the image super-resolution reconstruction method based on deep convolutional neural networks from three aspects: mathematical foundation and model constraints, feature reuse mechanisms, and multi-scale reconstruction strategies. In the mathematical foundation

part, the paper analyzes the linear system model of image degradation and the root cause of the ill-posedness of the inverse problem, clarifies the translation equivariance and frequency-domain filtering characteristics of convolution mapping, and discusses the trade-off path between the local receptive field and global dependencies. In the feature reuse part, the paper studies the enhancement effect of residual paths on gradient preservation and identity mapping convergence, analyzes the advantages of dense connections in cross-layer feature reuse and parameter efficiency optimization, and introduces the channel attention mechanism to achieve adaptive feature recalibration and frequency decomposition. In the reconstruction strategy part, the paper discusses the hierarchical feature decomposition and fusion framework guided by the Laplacian pyramid, elaborates the periodic shuffling principle of sub-pixel convolution, and explains the error iterative correction and reconstruction consistency constraint in the recursive back-projection architecture. The above research provides a theoretical basis for designing high-performance, low-redundancy super-resolution reconstruction networks. Future research directions include: exploring lightweight network structures for mobile deployment; studying unsupervised or self-supervised learning paradigms to reduce the reliance on paired training data; introducing multi-modal information to assist reconstruction and enhance detail recovery capabilities in complex scenes; and combining generative frameworks such as neural radiance fields or diffusion models to explore universal super-resolution reconstruction methods for arbitrary scaling factors and unknown degradation models.

References

[1] Nie Yalin, et al. "Super-Resolution Reconstruction of Open-Pit Mine Remote Sensing Images Fusing Deep Convolutional Neural Network and Swin Transformer." Metal Mine, no.12(2024):240-245.

[2] Zhang Jindi. Research on Single 3D-MRI Image Super-Resolution Reconstruction Based on Deep Convolutional Neural Network. 2024. Chongqing Medical University, MA thesis.

[3] Yuan Xilin, et al. "Infrared Image Super-Resolution Reconstruction Technology Based on Deep Convolutional Neural Network." Infrared Technology, vol.45, no.05(2023):498-505.

[4] Ni Ruoting, and Zhou Lianying. "Face Image Super-Resolution Reconstruction Method Based on Convolutional Neural Network." Computer and Digital Engineering, vol.50, no.01(2022):195-200.

[5] Xie Chao, and Zhu Hongyu. "Image Super-Resolution Reconstruction Method Based on Deep Convolutional Neural Network." Transducer and Microsystem Technologies, vol.39, no.09(2020):142-145.

[6] Fan Peipei. Single Depth Image Super-Resolution Reconstruction Based on Convolutional Neural Network. 2020. Xihua University, MA thesis.