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: *Image super-resolution reconstruction is an ill-posed inverse problem, so a good model of the natural image prior needs to be used to solve it. Deep Convolutional Neural Networks can be used to solve the problems of traditional hand-crafted features in this paper. Systematically study the theoretical foundation of this way, analyze the degradation linear model and the ill-posedness of the inverse problem at the level of mathematical foundation, clarify the transformation characteristics of convolution mapping and the constraints of receptive fields, explore the gradient preservation of residual paths, the feature reuse of dense connections, and the recalibration mechanism driven by channel attention at the level of feature reuse, and elaborate the Laplacian pyramid decomposition, the sub-pixel convolution shuffling operation, and the recursive back-projection error correction method at the level of reconstruction strategy. The three-dimensional framework of the theoretical basis for deep convolutional neural networks in super-resolution mentioned above consists of network structure, 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

Recently, super-resolution reconstruction of images has been used to solve many problems in remote sensing and medical diagnosis. In fact, it is an ill-posed inverse problem that needs to be regularised by adding prior information. The previous interpolation and sparse coding methods have the problem of hand-crafted features and are prone to edge blurring and ringing. Deep Convolutional Neural Networks have been trained to learn all the statistical laws of natural images, and now they combine feature extraction and reconstruction mapping in a single module. The reasons for using the above method are as follows: The depth and range of the network determine the receptive field and the extent of feature abstraction; therefore, it is necessary to build reasonable connections to avoid gradient attenuation; and the upsampling strategy should be both accurate and efficient, as a single feedforward structure cannot handle multi-scale information. In order to make systematic improvements in the behaviour of deep convolutional neural networks, the three problems of mathematical constraints, feature reuse and multi-scale reconstruction need to be solved to improve reconstruction accuracy and reduce model redundancy.

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

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

Generally speaking, the process of image degradation can be modelled by a linear forward model; that is to say, a high-resolution image is convolved with a blur kernel to obtain a low-resolution observation image, and then noise is added after downsampling. The low-resolution image can be expressed as the product of the degradation matrix and the high-resolution image vector plus a noise term, and this degradation matrix is the combined effect 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 has a serious column-underdetermined property, its null space is non-trivial, and

therefore there is no unique solution for the inverse mapping. The three attributes of ill-posedness are: the existence of a solution is sensitive to noise; non-uniqueness of the solution cannot be guaranteed due to loss of information; and the stability of the solution is very poor in the presence of observation errors because of a rapid decay of the singular values of the degradation matrix.

Add a prior to regularise the objective function and thus deal with the ill-posedness of the inverse problem and restrict the solution space. Functional analysis can be done in the form of implicit regularization; that is to say, the structure and parameterisation of the network map the manifold of high-resolution images into a low-dimensional feature space, and the local connectivity and weight-sharing properties of convolutional layers naturally limit the size of the hypothesis space. A non-linear activation function is used to introduce sparsity, and thus the reconstruction map can be expressed in terms of the statistics of natural images. Implicit Regularisation does not need to be added as a prior term; instead, it will be learned during the training of the end-to-end model, and thus the ill-posed nature of the inverse problem can be resolved[1].

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

A weighted sum of the neighbouring pixels is taken in a small window, and then the whole process is repeated by sliding the window across the image to create a multi-dimensional tensor in the feature space. The above mapping has translation equivariance; that is to say, if the input image is shifted, the corresponding shift in the output feature maps will be the same, and this is due to weight sharing in the convolution kernel. The Fourier Transform of the convolution kernel in the frequency domain shows that it is a filter. Convolution kernels of different sizes and parameters can be used to perform low-pass filtering and keep the structure of the image, or high-pass filtering can be used to obtain edges and details. A cascade of several convolutional layers is equivalent to a gradual division of the signal in feature space, and the feature maps obtained at each stage are the responses of the original image to different scales and orientations.

As the depth of the network increases, the feature space moves away from the geometric structure of the original pixel domain and a hierarchy of abstractions is formed. Shallow feature maps are not strongly correlated with pixels; rather, they contain relatively little information about edges, corners and simple textures, and thus their transformation properties are close to those of steerable filter banks. Group the local patterns in the deep feature map according to the different nonlinear operations, and as the spatial resolution decreases, the number of channels increases; thus, the type of representation is changed from low-order statistics to high-order structural descriptions. However, there will be some loss, and some high-frequency details will not be obtained in the layer-by-layer mapping. Cross-layer connections can directly pass the shallow features to the deep layers through skip connections; thus, there will be no loss of fine-grained information in the shift of feature space, and the multi-scale information required for the reconstruction task will be retained[2].

1.3 Constraints of Local Receptive Fields and the Trade-off with Global Context Dependence in Super-resolution Reconstruction

The Receptive Field of each output neuron in a standard Convolutional Neural Network is the product of the size of the convolution kernel and the number of network layers. To solve the problem of super-resolution reconstruction, the value of a single pixel can be obtained from both the local texture in its vicinity and the self-similarity pattern of distant areas in the image. A small receptive field in the local area cannot cover many repetitions of a pattern or the whole shape of a long edge, so there will be reconstruction errors or missing parts in the structure. If the receptive field is smaller than the support range of the degradation blur kernel, the model will be unable to carry out deconvolution properly and will thus increase the reconstruction error.

Dilated convolutions are used to increase the receptive field exponentially and reduce the number of parameters; thus, they can solve the problems of locality and non-locality. Dilated convolution has a gap between the elements of the convolution kernel to cover a larger area in the input and keeps the resolution of the output feature map the same as that of the kernel. A large number of dilations can be used to form a parallel structure of multi-scale receptive fields, and thus both local details and global layout information can be obtained by the network. Another way is to add a self-attention mechanism to get the correlation weights among all pairs of positions in the feature map and explicitly model non-local relationships. The above way can reduce the problem of repeated artifacts and discontinuous

boundaries in the reconstructed image. However, it should be pointed out that dilated convolution can cause grid artifacts, and self-attention has a quadratic computational complexity. Group dilation or sparse attention can be used to meet the demands of locality and globality more reasonably, and they are relatively efficient.

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

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

Deep Convolutional Networks will probably have the problem of vanishing or exploding gradients in backpropagation. Residual Paths are added to change how information is carried in the traditional sequential connection. The residual block directly passes the input through the identity mapping of the nonlinear transformation layers and then adds it to the output; thus, the gradient can be propagated directly from the deep layer to the shallow layer without going through the nonlinear activation functions and convolutions of each layer. The shortcut connection is a continuous gradient path and does not need to be extended. Optimisation theory: The residual structure changes the mapping so that what needs to be learned is the difference between the original direct fit and a perturbation fit of the identity mapping; that is, one learns the error. When the optimal mapping is close to the identity map, the residual branch only needs to learn a zero map, and it is much simpler than learning the identity map from scratch[3].

To keep the identity mapping, element-wise addition is performed between the skip connection and the output of the main path; it must be that the feature maps of the two branches have the same spatial size and number of channels. To reduce the number of features, linear projection or zero-padding can be used to modify the identity path, but linear projection will increase the number of parameters and directly affect the flow of gradients. The order of placing Batch Normalisation and activation functions in Deep Residual Networks affects the purity of the identity mapping. The pre-activation residual unit puts Batch Normalisation and the activation function before the convolution layer, so the identity path is not blocked by non-linearity and the gradient flow is relatively good. The above Design can guarantee good convergence of the network even for a depth of several hundred layers, and thus it is possible to train large-capacity models for super-resolution reconstruction.

2.2 Cross-layer Feature Reusage and Parameter-Efficient Optimisation of Dense Connection Structures

A dense connection network takes the output of each layer as the input for all the following layers, and to each layer, the set of feature maps from all the previous layers is given, and then a concatenation operation is performed along the channel dimension. The connection mode above can achieve explicit cross-layer feature reuse in the network; that is to say, the edge and texture information obtained in the shallow layers can be directly passed to the deep layers without attenuation at each level in residual learning. Dense connections are more parameter-efficient; that is to say, they can reduce the number of new feature maps added per layer by having a small growth rate, and thus the total number of parameters will be much smaller than that of a conventional convolutional network of the same depth. Each layer only needs to learn a small number of new features, and the rest of the features can be used in the computation results of the previous layers; thus, it is not necessary to extract the same feature repeatedly in all the layers.

A dense connection structure can gather all levels of information on the problem of super-resolution reconstruction, such as detailed information and general shape. The shallow feature maps have rich high-frequency edge information, the middle-level feature maps contain texture patterns, and the deep feature maps have semantic layout information. The good link will be employed to build the reconstruction at this time. However, the dense connection also has a storage overhead problem; that is to say, the number of feature map channels increases linearly with the number of layers. To reduce the number of features in the compression stage and save memory, convolution and pooling are used, and at the same time, reuse efficiency is also maintained. The Connection Mode of the dense blocks will also affect how well the features are reused. To reduce the number of channels in each dense block and thus lower the computational cost, add a bottleneck but keep the function of cross-layer information transmission.

2.3 Adaptive Feature Recalibration and Frequency Decomposition Guided by Channel Attention

All the feature channels have different purposes in the super-resolution reconstruction; some channels are used to carry low-frequency structural information, and others are employed to recover high-frequency details. Channel Attention Mechanism reduces the spatial information in a channel to a single value by global average pooling, and this value is usually related to the activation strength of that channel. Next, a two-layer non-linear transformation is employed in the gating mechanism to obtain a channel weight vector, and then this vector is used to weight the original feature maps along the channel dimension. Therefore, in order to increase the number of features, the weight of the channel that contributes more to reconstruction is amplified, and the activation of irrelevant or weak channels is reduced[4].

The division of frequency bands in the reconstruction stage can be done more reasonably with channel attention. The low-frequency part contains the general brightness and structure of the image, has a relatively smooth distribution in the feature map, and usually has high and stable channel weights; the high-frequency part is related to edges and fine details, is sparse and localised, and needs more detailed weight modulation. The Channel Attention Mechanism can learn to differentiate among different frequency areas indirectly, and by modifying the activation threshold of each channel, it is possible to determine how much of the high-frequency information needs to be kept. Add a multi-branch frequency decomposition module to the deep network to separate the input features into low-frequency approximation components and high-frequency residual components, and then send both of them to the attention module separately. The low-frequency branch has a large receptive field and can learn general structural information; the high-frequency branch is more detailed in its attention weights and is thus better at detecting fine-grained information accurately. Frequency-aware attention can be added to modify the features of the network based on the frequency distribution of the input image, and it is hoped that both the objective and subjective evaluation results of reconstruction accuracy will improve.

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

3.1 Hierarchical Feature Decomposition and Fusion Guided by the Laplacian Pyramid

The residual components at all levels of spatial frequency in the Laplacian pyramid are obtained by successive downsampling and then upsampling and taking the difference. All the levels of the pyramid have different scales, and the top level contains the low-frequency approximation part of the image, and the lower levels contain high-frequency detail residuals. Hierarchical feature decomposition based on Laplacian pyramids is used to divide the reconstruction problem into several parts, so the low-frequency component is employed to reconstruct structures with a large receptive field, and shallow sub-networks are applied to predict the high-frequency components. To prevent spectral aliasing in the processing of multi-band data over a single network, decomposition can be employed[5].

Top-down hierarchical feature fusion is carried out at the various levels of the pyramid. Upsample the high-level low-frequency feature, and then either concatenate or add it to the residual feature of the next level to get a fused feature map. Repeat the above fusion process several times to reach the bottom of the pyramid at the original image resolution. All the scales of a Laplacian pyramid correspond to a hierarchy of features extracted by convolutional networks; that is to say, shallow feature maps are high-frequency residual components, and deep feature maps are low-frequency structural components. Design cross-layer connections to map the internal features of the network to each level in the pyramid, achieve end-to-end joint optimisation, have the network automatically learn the best residual representation at different scales, and thus improve the scale consistency of the reconstruction results.

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

The first transposed convolution has a problem of checkerboard artifacts during upsampling, and the reason for this problem is an uneven overlap pattern of the convolution kernel. Periodically shuffle to avoid this problem; that is, the sub-pixel convolutional layer performs standard convolution on the low-resolution feature map and outputs an intermediate feature map with a number of channels equal to the square of the upsampling factor, and then it rearranges the information in the channel dimension to the spatial dimension by pixel shuffling. The main idea of periodic shuffling is to concentrate the

convolution computation in the low-resolution space and then increase the resolution via a fixed rearrangement map, so that the upsampling process does not add redundant learnable parameters[6].

Based on the theory of information, the upsampling of the low-resolution feature map in the sub-pixel convolutional layer to obtain a high-resolution one has periodic local dependence in its mapping function. The value of each high-resolution pixel is obtained by combining the information in several channels around it in the low-resolution area, and these channels carry information for different sub-pixel positions before being organised. Periodical Shuffling is used to keep the spatial continuity of the feature map and to prevent periodic artifacts caused by zero-padding in transposed convolution. Generally speaking, in a super-resolution reconstruction network, the sub-pixel convolutional layer is placed at the end of the network, and it is given the multi-channel feature maps obtained by the deep feature extraction module. By changing the relationship between the number of channels of the intermediate feature map and the upsampling factor, it is possible to increase the spatial resolution by an integer multiple or a non-integer multiple; however, one must also consider the trade-off among the visual quality of the reconstructed results and computational cost.

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

Recursively back-projection is based on the error feedback method of the iterative back-projection algorithm; that is, repeatedly carry out an up-projection operation and a down-projection operation to reduce the reconstruction error gradually. Up-projection reduces the current estimation of the high-resolution image to get a low-resolution residual, and then down-projection back-projects this residual in the high-resolution space to update the estimate. Back-projection is carried out, a correction term is calculated, and then this correction term is added to the current estimate to obtain a new, high-resolution image. Recursion increases the number of iterations to that of a network module sharing weights, and each module carries out a complete back-projection update; these modules are connected sequentially to form a recursive chain.

The consistency constraint of reconstruction is that if the same degradation process is applied to the finally reconstructed high-resolution image, it should result in the original low-resolution image. Recursively back-projection computes the difference between the degraded low-resolution image and the original input at each step, and then this difference is taken as the error signal for the next correction. An internal consistency check is carried out above to confirm that the reconstructed result is in agreement with the degradation model; otherwise, it will be one of the typical problems of traditional feedforward networks, namely texture drift or structural deformation. The depth of recursion is the number of error corrections; a deeper recursive structure can achieve better accuracy, but it should be noted that the gradient may be reduced or become very large in the recursive chain. Residual connections can be added to directly connect the input of each back-projection module to its output to stabilise the recursive training process and keep the progressive approximation property of error correction.

Conclusion

The three parts of this paper are: (1) Mathematical Foundation and Model Constraints of Deep Convolutional Neural Network-based Image Super-Resolution Reconstruction; (2) Feature Reuse Mechanism; and (3) Multi-scale Reconstruction Strategy. In the section of mathematical foundations, we will present the linear system model of image degradation and analyze the root cause of ill-posedness in the inverse problem; we will clarify the translation equivariance and frequency-domain filtering properties of convolution mapping; and we will discuss the trade-off path between the local receptive field and global dependency. In the part about feature reuse, we will discuss how to improve the effect of residual paths in terms of gradient preservation and convergence of the identity mapping, study the advantages of dense connections for cross-layer feature reuse and parameter-efficient optimization, introduce a channel attention mechanism for adaptive feature recalibration and frequency decomposition, etc. In the part of Reconstruction Strategy, we introduce the hierarchy of feature decomposition and fusion based on the Laplacian pyramid, and also present the periodic shuffling principle of sub-pixel convolution and the error iterative correction and reconstruction consistency constraint in the recursive back-projection architecture. In recent years, some scholars have conducted studies on the Design of high-performance, low-redundancy super-resolution reconstruction networks. The first direction of research going forward will be the construction of a lightweight network for

mobile applications. Unsupervised or self-supervised learning will be employed to reduce the requirement for paired training data, various types of information will be integrated to enhance the reconstruction results and improve fine details in complex scenes, and general super-resolution reconstruction methods will be developed under all kinds of scaling factors and unknown degradation models by combining different generative frameworks, such as neural radiance fields and diffusion 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.