Single Image Super-Resolution Using Densely Connected Convolutional Networks and Inception-ResNet

Chao Chen

School of Business, Shandong Normal University, Jinan 250000, China.

1192613826@qq.com

Abstract

Recent studies have shown that the use of deep convolutional neural networks (Deep CNNs) can significantly improve the performance of single-image super-resolution reconstruction. Although the existing super-resolution convolutional neural networks have achieved good high-resolution image reconstruction by deepening the number of network layers, it also makes these networks harder to train and easier to overfit. In this paper, we propose a highly accurate and fast single-image super-resolution reconstruction (SISR) method by introducing densely connected convolutional networks and Inception-ResNet(DCIRSR). To enhance feature propagation and facilitate feature reuse, the feature extraction network in the model first uses feedforward to connect each layer to other layers in the network, and then uses all previous layer feature maps as inputs for each subsequent layer. In the feature reconstruction network, the parallelized 1x1 CNN and 3x3 CNN structures reduce the size of the previous layer of output, enabling faster calculations while reducing information loss and allowing direct processing of the original image. In addition, we only learn the residuals between the high-resolution images and the low-resolution images. Results on benchmark datasets demonstrate that the proposed model not only achieves higher accuracy but also enables faster and more efficient calculations than the state-of-the-art methods.

Keywords

Deep Learning, Image Super Resolution, Deep CNNs, Dense connection, Inception-ResNet.

1. Introduction

Resolution is an important indicator for evaluating image quality. High-resolution images have a high pixel density and contain more details. These details have important application values in the fields of monitoring equipment, satellite imagery, and medical imaging. The resolution depends on the effects of the imaging system and other factors. Single Image Super-Resolution (SISR) is a technique that overcomes the inherent limitations of imaging hardware such as image sensors and improves the image quality. It uses computer software to reconstruct high resolution (HR) images from a low resolution (LR) image. Recently, methods based on deep convolutional neural networks have achieved tremendous performance improvements in the problem of SISR from LR to HR.

With the development of convolutional neural networks, its architecture has become deeper and more complex, which has led to a new problem that arises as its performance increases: When input or gradient information reaches the end of networks through many layers, it will disappear and "wash out". There was some very good works to solve this problem. A data-bypass (skip-layer) technique has been proposed in the Highway Networks [1] and ResNet [2] architectures to allow signals to flow at high speed between the input layer and the output layer. The core idea is to create a cross-layer connection to connect to the Internet. Before and after the layer, the signal is passed from one layer to the next. In order to maximize the flow of information between all layers in the network, DenseNet [3] connects any two layers in the network, so that each layer in the network accepts the features of all layers in front of it as input. In addition, deeper neural networks are also more difficult to train. He et al. proposed a residual learning framework(ResNet) to ease network training. Unlike other networks that learn from unreferenced functions, the residual network explicitly defines the layer as the reference input layer to learn the residual function. The proposal of ResNet solves the problem that the previous network structure could not be trained when it is deep and can improve the accuracy by significantly increasing the depth of the network. Residual network architecture has been used in a lot of works.

Inspired by the above deep learning model, this paper proposes an image super resolution reconstruction algorithm based on densely connected convolutional networks networks and Inception-ResNet [4]. The algorithm mainly includes the following features: firstly, directly use convolutional neural network to establish the end-to-end mapping model without preprocessing the image; secondly, using skip connection to connect the local information on the image with the global information, it can also be said to use the advanced features and the underlying features to reconstruct the image. Thirdly, using residual learning and in-depth network structure to effectively improve the ability of network learning while reducing the training time. Lastly,using ADAM to replace the traditional SGD optimization method, and reduce the learning rate during the training process, and the algorithm performance improvement effect is obvious.

2. Related Work

Deep Learning-based methods are currently active and showing significant performances on SISR tasks. In 2014, Dong et al. [5] proposed a CNN-based image super-resolution reconstruction method—Super-Resolution Convolutional Neural Network (SRCNN),which is the pioneering work of deep learning for super-resolution reconstruction. This method can learn end-to-end mapping from low resolution images to high resolution images. The SRCNN first uses a bicubic interpolation to amplify the low resolution image to the target size, then fits the nonlinear map through a three-layer convolution network, and finally outputs the high resolution image results. Later, Dong et al. [6] proposed FSRCNN, which is an improved version of SRCNN for Image Super-Resolution. There are three main aspects: First, at the end, a deconvolution layer is used to enlarge the size, so the original low-resolution image can be directly input into the network, instead of being enlarged by the bicubic method as in the previous SRCNN. The second is to change the feature dimension, use a smaller convolution kernel and use more mapping layers. The third is to share the mapping layer. If we need to train different upsampling models, we only need fine-tuning the final deconvolution layer..

In 2016, DRCN[7] was proposed by Jiwon Kim et al. ,whicn applies a Recursive Neural Network (RNN) structure to super-resolution processing for the first time. At the same time, using the Skip-Connection idea deepened the network structure (16 recursive), increased the network receptive field, and achieved significant performances. Inspired by VGG-Net [8] and ResNet, the same authors of DRCN proposed another high-precision SISR method based on Deep CNNs—VDSR [9]. VDSR can train different proportions of sub-images in the same batch , and use residual learning and adjustable gradient cuts to improve the convergence speed. Inspired by ResNet, VDSR and DRCN, DRRN [10] proposed by Tai Y et al. adopted a deeper network structure to obtain image super-resolution reconstruction performance.

Under the premise of ensuring a small change in accuracy, the image super resolution reconstruction models based on deep learning are also developing in a shallower and faster direction. The running time of RAISR [11] proposed by Yaniv Romano et al. and DCSCN [12] proposed by Jin Yamanaka et al. is one to two orders of magnitude faster than other state-of-the-art methods based on deep learning, and the results are sometimes even better than the existing technology, such as DRCN, VDSR or RED.

3. Proposed Method

3.1 Model Overview

As shown in Fig. 1, our model (DCIRSR) consists of two parts: feature extraction network and feature reconstruction network. We directly use the original image as input, extract the local and global features of the image through the dense module, and then import the extracted features into the parallel CNN network to reconstruct the image details. The input low resolution image and the output high resolution image are largely similar. In other words, the low frequency information carried by the low resolution image is similar to the low frequency information of the high resolution image.

Training these low frequency information will take a lot of time. In fact, we only need to learn the high-frequency partial residual between the high-resolution image and the low-resolution image. It is easier to learn the residuals with deep networks. For networks of the same depth, the residual network converges faster and is easier to optimize.

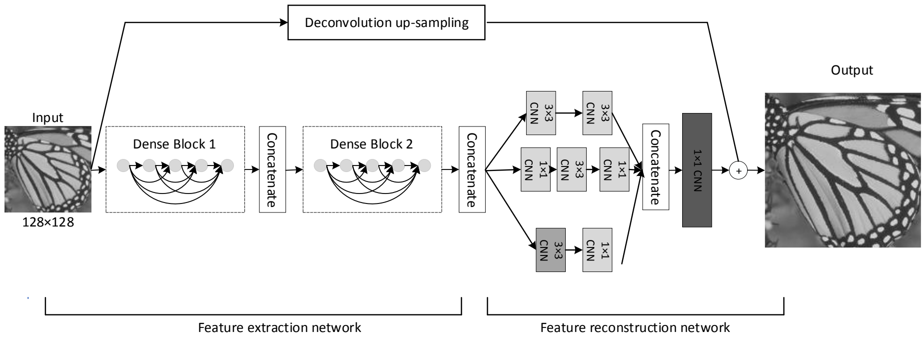


Fig 1. Our model (DCIRSR) structure.

3.2 Feature Extraction Network

For image super-resolution reconstruction tasks, the more features extracted, the better the reconstruction effect. Previous image super-resolution reconstruction models typically ignored the role of underlying low-level features that could provide additional information for reconstructing high-frequency details in HR images, while only using top-level advanced functionality to reconstruct HR images. Unlike previous work, we use dense skip joins to create short paths from low-level features to high-level features to get more contextual information. This not only helps ensure that the largest amount of information and mitigation gradients disappear in the network layer, but also makes training easier.

In the feature extraction network, we use a set of DenseNet blocks to learn the relevant features. The DenseNet structure was first proposed in [3]. Unlike ResNets proposed in [2], each layer in DenseNet gets additional input from all previous layers and passes its own feature map to all subsequent layers instead of merging features from all previous layers to the last layer, see Fig. 2. Therefore, the i^{th} layer takes the feature maps of all previous convolutional layers as input:

$$x_i = H_i([x_0, x_1, \cdots, x_{i-1}]) \tag{1}$$

Where $[x_0, x_1, \cdots, x_{i-1}]$ represents the connection of feature maps generated in the convolutional layers $1, 2, \cdots, i-1$ before the i^{th} layer, $H_i(\cdot)$ can be regarded as a composite function of two consecutive operations: 1×1 CNN or 3×3 CNN, followed by Parametric ReLU.

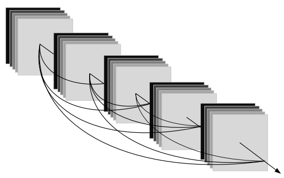


Fig 2. A 5-layer dense block with a growth rate of k = 4.Each layer takes all preceding feature-maps as input.

3.3 Image Reconstruction Network

Inspired by GoogLet [13] and ResNet, we use the parallel CNN structure and residual learning in the image reconstruction network so that the network can achieve very good performance at a relatively low computational cost, and at the same time, it can significantly accelerate the initial network training.

Filters with smaller size have better performance than larger filters with the same input size and output depth. Therefore, we use 1x1 CNN and 3x3 CNN in parallel networks. Since all features generated by the feature extraction network are cascaded before the input layer of the image reconstruction network, the input data of the latter will have a larger dimension. Therefore, we use 1x1 CNN to reduce the size of the input layer. At the same time, it also increases the nonlinearity of the network.The parallel CNN structure is followed by a 1x1 CNN for the dimensional transformation of the filter bank, thus compensating for the dimensional reduction caused by the parallel CNN structure. The DCIRSR reshapes the LR images processed by the feature extraction network and the parallel CNN structure into HR images. Our model only learn the residual error between the low-resolution image after deconvolution interpolation and the corresponding high-resolution image. This will not only significantly increase the speed of network training, but also bring about an improvement in image reconstruction performance.

4. Experiments

4.1 Datasets for Training and Testing

The experiment used 391 different color images as the training dataset, which are 91 images from Yang et al. [14] and BSD100 and BSD200 from the Berkeley segmented dataset [15].To expand the training data set, we flipped each image horizontally or vertically and cropped into smaller images. In the test phase, Set5 [16] and Set14[17] are used as test datasets. In order to compare with the existing image super-resolution algorithm, this paper convert color (RGB) images to YCbCr images. This article only trains and tests the Y channel of the image.

4.2 Training Setup

In DCIRSR model, the activation function of each layer adopts Parametric Rectified Linear Unit (PReLU) [18], which can be regarded as ReLU activation function with correction parameters. Compared to the ReLU activation function, the PReLU activation function only adds a small amount of computation to achieve a higher accuracy and can avoid the "dying ReLU" phenomenon caused by ReLU. At the same time, we use the method proposed by He et al. to initialize the weights of each layer.

Mean Squared Error (MSE) of the output reconstructed image and the high-resolution surface image as a loss function. To prevent over-fitting problems, we added L2 regularization that describes the complexity of the model in the loss function. In general, model complexity is determined only by weights. In order to reduce the consumption of computing resources and make the model faster convergence, we use Adam [19] instead of the traditional SGD optimization method to minimize the loss function and reduce the learning rate during training where α=0.002, β1=0.9, β2=0.999 and ε=10E−8. Compared with other adaptive learning rate algorithms, Adam has faster convergence, more effective learning, and can correct problems in other optimization techniques such as disappearance of learning rate and slow convergence.

4.3 Comparisons with State-of-the-Art Methods

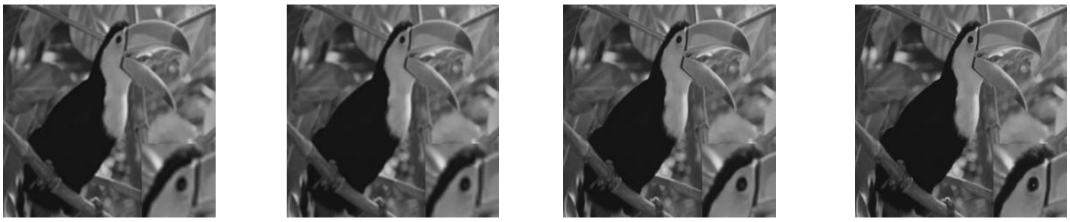
The existing image super-resolution evaluation criteria include subjective evaluation and objective quantitative evaluation. Since the subjective evaluation is to evaluate the image quality by the human eye observation output image, there is a great uncertainty. Therefore, we use Peak Signal to Noise Ratio (PSNR) which is currently the most common objective quantitative evaluation method to compare the accuracy of the proposed DCIRSR and other SR algorithms based on deep learning in experiments. The higher the PSNR value (in dB) between the two images, the less the distortion of the reconstructed image relative to the high-resolution image. Table 1 shows the comparison of the average PSNR (dB) between our model and other image super resolution reconstruction models for SISR.

Table 1. Comparison of Average PSNR (db) of Different Image Super-resolution Models

Dataset	scale	Bicubic	SRCNN	VDSR	DCIRSR (ours)
Set 5	×2	33.66	36.34	37.53	37.78
	×3	30.39	32.39	33.66	33.97
	×4	28.42	30.09	31.35	31.71
Set14	×2	30.24	32.18	33.03	33.24
	×3	27.55	29.00	29.77	29.93
	×4	26.00	27.20	28.01	28.22
BSDS100	×2	29.56	30.71	31.90	32.01
	×3	27.21	28.10	28.82	28.89
	×4	25.96	26.66	27.29	27.36

We compared the proposed DCIRSR with the existing image super-resolution algorithm, including Bicubic, SRCNN and VDSR, each of which magnifies the image by 2, 3, and 4 times. As shown in the above experimental results, the PSNR of the proposed algorithm is better than other algorithms, which shows that DCIRSR has excellent image reconstruction performance.

The image used for testing comes from Set5 . Fig. 3 shows the visual effects of an image before and after DCIRSR reconstruction. In order to make the visual effect more obvious, we took the same size area from the same position in each picture and then enlarged it twice (as shown in the lower right corner of each image). Compared with bicubic, the reconstructed image of our algorithm has the best visual effect, which is the closest to the ground truth image.



Input bicubic result ground truth

Fig 3. An example of our result of img_002 in set5.

5. Conclusion

This paper proposed an image super-resolution algorithm based on densely connected convolutional networks and Inception-ResNet. The model (DCIRSR) reconstructs high resolution (HR) images from a low resolution (LR) image through two branches: one branch directly inputs the low-resolution image processed through deconvolution interpolation to the last layer of the network; the other branch reconstructs a high-resolution image through a low- resolution image, which consists of a feature extraction network and a feature reconstruction network. The feature extraction network extracts local and global image features through density blocks, and the feature reconstruction network reconstructs the extracted features into high resolution images through Inception-ResNet. Our model learns the residual part between the low-resolution image after the deconvolution interpolation and the corresponding high-resolution image by the feature extraction network and the feature reconstruction network, so that the training speed is faster and the optimization is easier. The experimental result of image super-resolution reconstruction shows that the model has better reconstruction performance than other deep learning super-resolution reconstruction methods.

Acknowledgements

This work was supported by the Natural Science Foundation of China (No.61502283). Natural Science Foundation of China (No.61472231). Natural Science Foundation of China (No.61640201).

References

[1] Srivastava R K, Greff K, Schmidhuber J. Training very deep networks[C]//Advances in neural information processing systems. 2015: 2377-2385.
[2] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition [C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.
[3] Huang G, Liu Z, Weinberger K Q, et al. Densely connected convolutional networks [C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017, 1(2): 3.
[4] Szegedy C, Ioffe S, Vanhoucke V, et al. Inception-v4, inception-resnet and the impact of residual connections on learning[C]//AAAI. 2017, 4: 12.
[5] Dong C, Loy C C, He K, et al. Learning a deep convolutional network for image super-resolution [C]//European Conference on Computer Vision. Springer, Cham, 2014: 184-199.
[6] Dong C, Loy C C, Tang X. Accelerating the super-resolution convolutional neural network [C]//European Conference on Computer Vision. Springer, Cham, 2016: 391-407.
[7] Kim J, Kwon Lee J, Mu Lee K. Deeply-recursive convolutional network for image super-resolution [C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 1637-1645.
[8] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition [J]. arXiv preprint arXiv:1409.1556, 2014.
[9] Kim J, Kwon Lee J, Mu Lee K. Accurate image super-resolution using very deep convolutional networks [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016: 1646-1654.
[10] Tai Y, Yang J, Liu X. Image super-resolution via deep recursive residual network [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017, 1(2): 5.
[11] Romano Y, Isidoro J, Milanfar P. RAISR: rapid and accurate image super resolution [J]. IEEE Transactions on Computational Imaging, 2017, 3(1): 110-125.
[12] Yamanaka J, Kuwashima S, Kurita T. Fast and Accurate Image Super Resolution by Deep CNN with Skip Connection and Network in Network [C]//International Conference on Neural Information Processing. Springer, Cham, 2017: 217-225.
[13] Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions [C]. Cvpr, 2015.
[14] Yang J, Wright J, Huang T S, et al. Image super-resolution via sparse representation[J]. IEEE transactions on image processing, 2010, 19(11): 2861-2873.
[15] Arbelaez P, Maire M, Fowlkes C, et al. Contour detection and hierarchical image segmentation [J]. IEEE transactions on pattern analysis and machine intelligence, 2011, 33(5): 898-916.
[16] Bevilacqua M, Roumy A, Guillemot C, et al. Low-complexity single-image super-resolution based on nonnegative neighbor embedding [J]. 2012.
[17] Zeyde R, Elad M, Protter M. On single image scale-up using sparse-representations [C]//International conference on curves and surfaces. Springer, Berlin, Heidelberg, 2010: 711-730.
[18] Xu C, Liu T, Tao D, et al. Local rademacher complexity for multi-label learning [J]. IEEE Transactions on Image Processing, 2016, 25(3): 1495-1507.
[19] Kingma D P, Ba J. Adam: A method for stochastic optimization [J]. arXiv preprint arXiv:1412.6980, 2014.