2018 9th International Conference on Information Technology in Medicine and Education

Single Image Super-Resolution Using Deep CNN with Dense Skip Connections and Inception-ResNet

Chao Chen 1st
Management Science and Engineering
Shandong Normal University
Jinan, China
1192613826@qq.com

Feng Qi 2nd
Management Science and Engineering
Shandong Normal University
Jinan, China
cliff@sdnu.edu.cn

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. In this paper, we propose a highly accurate and fast single-image super-resolution reconstruction (SISR) method by introducing dense skip connections and Inception-ResNet in deep convolutional neural networks. In the proposed network, all previous layer feature maps are used as input for each subsequent layer to promote feature reuse and alleviate the vanishing-gradient problem. In addition, the parallelized CNNs structure used reduces the size of the output of the previous layer, thereby accelerating the calculation speed and reducing the feature loss. Moreover, we only learn the residuals between the high-resolution images and the low-resolution images, and use the adjustable gradient cropping to achieve a very high learning rate. 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: Image Super Resolution, Deep Learning, Convolutional Neural Network, Skip Connection, Inception-ResNet

I. 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 dense skip connections 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.

II. RELATED WORK

Deep Learning-based methods are currently active and showing significant performances on SISR tasks. In 2014,

2474-3828/18/$31.00 ©2018 IEEE
DOI 10.1109/ITME.2018.00222

Dong et al. [5] applied CNNs to solve the SR problem for the first time and proposed a CNN-based image super-resolution reconstruction method—Super-Resolution Convolutional Neural Network (SRCNN). This method can learn end-to-end mapping from low resolution images to high resolution images. The SRCNN first uses bicubic interpolation to amplify the low-resolution image to a target size, then fits the nonlinear mapping through a three CNN layers, and finally outputs high-resolution image results. Later, Dong et al. [6] proposed FSRCNN, which is an improved version of SRCNN for Image Super-Resolution. On the one hand, the FSRCNN introduces a deconvolution layer at the end of the network for up-sampling, so that the original low-resolution image can be directly input into the network. On the other hand, the FSRCNN uses smaller filter size and deeper network structure. These improvements make FSRCNN achieves an acceleration of more than 40 times, and even has a more excellent reconstruction effect.

In 2016, Deeply-Recursive Convolutional Network for Image Super-Resolution (DRCN) [7] was proposed by Jiwon Kim et al. DRCN 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. The RED network [10] with a depth of 30 layers consisting of a symmetric convolutional-deconvolution layer also uses the residual network. The structure of the RED network is symmetrical, and each convolutional layer has a corresponding deconvolution layer. The convolutional layer is used to obtain the abstract content of the image, and the deconvolution layer is used to enlarge the feature size and restore the image details. As an encoding-decoding framework, it can learn end-to-end mapping from low-quality images to original images.

Ledig, Christian, et al. [11] proposed the SRGAN, which for the first time used the Generative Adversarial Network (GAN) to solve the problem of image super-resolution. Unlike other image super-resolution reconstruction methods, which use the mean square error as the loss function to pursue a high peak signal to noise ratio, SRGAN not only optimizes SRResNet (SRGAN's generation network part) using the mean square error to get a high peak signal to noise ratio, but also uses perceptual loss and adversarial loss to enhance the realism of the recovered picture.

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 [12] proposed by Yaniv Romano et al. and DCSCN [13] 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.

III. PROPOSED METHOD

A. Model Overview

As shown in Figure 1, our model (DSIRSR) is a fully convolutional neural network consisting of a feature extraction network and a 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. Because it is difficult to directly use the deep convolutional neural network to learn the potential identity mapping between the low-resolution image and the corresponding high-resolution image, our model learns the residual error between the low-resolution image after bicubic interpolation and the corresponding high-resolution image.

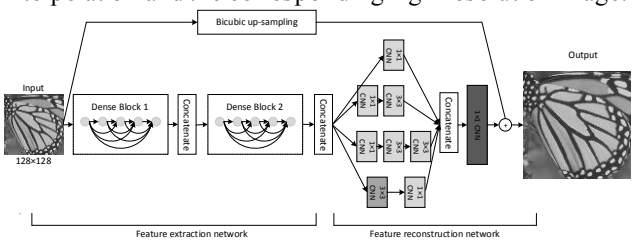

Figure 1. Our model (DSIRSR) structure.

B. Feature Extraction Network

Feature extraction plays an important role in image super-resolution reconstruction. The more features extracted, the better the reconstruction effect. Previous image super-resolution reconstruction models usually only use top-level advanced features for reconstructing HR images, while ignoring the role of underlying low-level features in SISR. SISR may benefit from different levels of features, and low-level features can provide additional information for reconstructing high-frequency details in HR images. In addition, reusing the underlying features helps to reduce feature redundancy and thus learn more compact CNN models. Unlike previous works, we used dense skip connections to combine low-level and high-level features during feature extraction to provide rich information for SISR. Dense skip connections obtain more contextual information by creating short paths from top level features to underlying features to better predict data in HR images. This not only helps to ensure that the maximum amount of information and mitigation gradients disappear across the network layers, but also makes it easier to train.

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. Therefore, the i^{th} layer takes the feature maps of all previous convolutional layers as input:

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

Where $[x_0, x_1, \cdots, x_{i-1}]$ represents the connection of feature maps generated in the convolutional layers $1, 2, \cdots, \mathrm{i}-1$ 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.

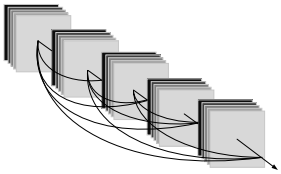

Figure 2. A 5-layer dense block with a growth rate of k = 4. Each layer takes all preceding feature-maps as input.

C. Image Reconstruction Network

GoogLeNet [14] uses an "Inception module" to connect feature maps generated by filters of different sizes. ResNet can improve accuracy by introducing a residual function that allows the network to increase depth significantly, while also making the network easier to optimize. Inspired by GoogLet 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.

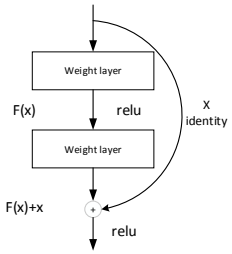

Figure 3. Residual learning: a building block.

On the one hand, spatial aggregation can be done on lower-dimensional embedding without causing any or any loss in presentation capabilities. On the other hand, 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. Because all the 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. Thus, we use 1x1 CNN to reduce the dimension of input layer where the computational requirements will increase too much. In other words, before the expensive 3×3 convolution, the 1×1 convolution product is used to calculate dimensionality reduction. At the same time, it also increases the depth of the network and 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 DSIRSR reshapes the LR images processed by the feature extraction network and the parallel CNN structure into HR images. Since the LR images processed by the bicubic interpolation and HR images are largely similar, we only use the residual connection to learn the high-frequency partial residuals between the two, which can significantly improve the speed of network training. It will also bring about an increase in network performance.

IV. EXPERIMENTS

A. Datasets for Training and Testing

The experiment used 391 different color images as the training dataset, which are 91 images from Yang et al. [15] and BSD100 and BSD200 from the Berkeley segmented dataset [16].

To expand the training data set, we flipped each image horizontally or vertically and cropped into smaller images. In the test phase, Set5 [17] and Set14[18] 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 (YCrCb color space, where the Y channel represents the brightness).

B. Training Setup

In DSIRSR model, the activation function of each layer adopts Parametric Rectified Linear Unit (PReLU) [19], 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 (Adaptive Moment Estimation) [20] instead of the traditional SGD (Stochastic Gradient Descent) 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 or the high variance of the parameter update causes the loss function to fluctuate.

C. 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 DSIRSR 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. Table1 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	DSIRSR (ours)
Set 5	×2	33.66	36.34	37.53	37.77
	×3	30.39	32.39	33.66	34.02
	×4	28.42	30.09	31.35	31.72
Set14	×2	30.24	32.18	33.03	33.18
	×3	27.55	29.00	29.77	29.91
	×4	26.00	27.20	28.01	28.23
BSDS100	×2	29.56	30.71	31.90	31.99
	×3	27.21	28.10	28.82	28.87
	×4	25.96	26.66	27.29	27.35

We compared the proposed DSIRSR 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 DSIRSR has excellent image reconstruction performance.

Input

bicubic

result

ground truth

Figure 4. An example of our result of img_002 in set5 .

Figure 4 shows the visual effects of an image before and after DSIRSR 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).

V. CONCLUSION AND FUTURE WORKS

This paper proposed an image super-resolution algorithm based on deep convolutional neural network. The model(DSIRSR) predicts the result of image super-resolution reconstruction through two branches: one branch directly inputs the low-resolution image processed through bi-cubic 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. Our model learns the residual part between the low-resolution image and the corresponding high-resolution image after the bicubic interpolation 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.

ACKNOWLEDGMENT

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] Mao X, Shen C, Yang Y B. Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections[C]//Advances in neural information processing systems. 2016: 2802-2810.

[11] Ledig C, Theis L, Huszár F, et al. Photo-realistic single image super-resolution using a generative adversarial network[J]. arXiv preprint, 2017.

[12] Romano Y, Isidoro J, Milanfar P. RAISR: rapid and accurate image super resolution[J]. IEEE Transactions on Computational Imaging, 2017, 3(1): 110-125.

[13] 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.

[14] Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions[C]. Cvpr, 2015.

[15] 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.

[16] 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.

[17] Bevilacqua M, Roumy A, Guillemot C, et al. Low-complexity single-image super-resolution based on nonnegative neighbor embedding[J]. 2012.

[18] 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.

[19] 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.

[20] Kingma D P, Ba J. Adam: A method for stochastic optimization[J]. arXiv preprint arXiv:1412.6980, 2014.