Piecewise-recursive Convolutional Network for Fast and Accurate Face Image Super-resolution

Pengyan Xie[1, a], Yanxiong Niu[1, b, *], Haisha Niu[1, 2, c], Dan Guo[1, d]

[1]*Institute of Instrument Science and Photoelectric Engineering, Beihang University, Beijing 100191, China*

[2]*Institute of instrument Science and Photoelectric Engineering, Beijing Information Science and Technology University, Beijing 100192, China*

[a]*xiepengyan@buaa.edu.cn,* [b]*niuyx@buaa.edu.cn,* [c]*niuhs@buaa.edu.cn,* [d]*2681470897@qq.com*

Keywords: face images, super-resolution, deep CNNs, recursive convolutional networks, skip connection, network

Abstract: Deep convolutional neural networks (Deep CNNs) have recently demonstrated high-quality reconstruction for face image super-resolution. However, as the depth grows, more computations are required and it is difficult to train the network. In this paper, a highly efficient and faster face image super-resolution method using a piecewise-recursive convolution network (PRCN) is proposed. Original low-resolution (LR) images are used as the inputs of the proposed model and thus significantly reduce the calculation cost. A combination of recursive convolutional networks and skip connection layers are used to extract both local and global features of input LR face images. Specially, the number of each recursive convolutional layer is optimized to further improve the performance and reduce the computation. For image reconstruction, 1×1 convolutional layers are used to reduce the dimension of the extracted features. Parallelized CNNs are then applied to learn an effective nonlinear mapping from the low-resolution (LR) to the high-resolution (HR) features. Experimental results show that the proposed algorithm outperforms the state-of-the-art methods, while achieving faster and more efficient computation.

1. Introduction

Face recognition technology has been widely used in in intelligent surveillance, identity authentication, human-computer interaction and digital entertainment. However, due to the limit of intrinsic device factors and the interference of external environmental factors, the obtained face images are usually of low resolution. More details are required when such images are applied in face recognition. Therefore, super-resolution (SR) technology is processed to transform low-resolution (LR) images into high-resolution (HR) images. During the process, the missing high frequency details are estimated.

Recently, Deep Learning (DL) models have been widely used in computer vision due to the powerful learning ability. Dong et al. [1] first proposed a deep learning-based SR method to predict the nonlinear LR-to-HR mapping. The model, which is called Super-Resolution Convolutional Neural Network (SRCNN), significantly outperforms classical non-DL methods. Based on SRCNN,

many deep CNN-based methods [2, 3, 4, 5, 6, 7] have been proposed. These methods outperform the previous shallow CNN-based methods by a large margin, which reflects the trend of 'the deeper the better' in SR.

Despite achieving excellent performance, deep CNNs require large computation and a lengthy processing time. To address this drawback, we propose a Piece-Recursive Convolutional Network (PRCN) in this paper. As shown in Fig. 1, the proposed method achieves state-of-the-art reconstruction performance with at least 10 times lower computational cost. Specifically, PRCN has two major algorithmic novelties:

(1) *Piecewise-recursive structure* is proposed in PRCN to keep the model compact. The recursive structure can enhance the potential representation of the network by increasing depth without adding additional parameters. However, with the increase of recursions, the network can be difficult to train due to the vanishing/exploding gradients problems. PRCN is relieved from this burden by introducing several recursive modules. Each module contains an ordinary convolutional layer and a recursive convolutional layer. The number of filters in each module is also optimized and thus results in better performance with faster computation.

(2) *A combination of skip connection and network in network* are introduced in PRCN. Since the local feature is more important than the global feature in SR, each output of ordinary and recursive convolutional layers is passed to the reconstruction network via skip connection. All these features are then concatenated as the input of the reconstruction network. A parallelized CNN structure [8] is used in PRCN. On the one hand, 1×1 convolutional layers can effectively reduce the input dimension and thus makes the network more concise. On the other hand, the parallelized structure can enhance the learning ability of the network at the cost of less computation compared with the chain structure.

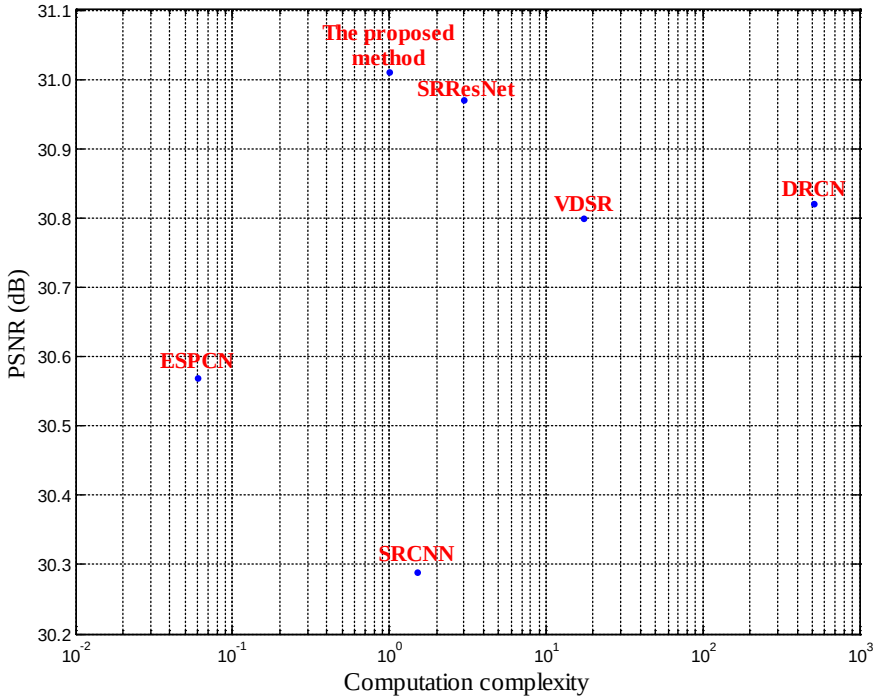

Fig. 1 Comparisons between reconstruction performance and computation complexity. The complexity of the proposed model is taken as 1.

2. Proposed Method

2.1 Model Overview

The proposed model, outlined in Fig. 2, consists of two sub-networks: feature extraction and image reconstruction networks. In the feature extraction network, multiple cascaded convolutional layers are used to extract the features of the input LR image. The extracted features are then connected to the reconstruction network as skip connection. In the image reconstruction network, parallelized 1×1 convolutional layers are used to reduce the dimension of concatenated features. The last convolutional layer outputs the final LR feature maps of 16 channels (when the scale factor s = 4). These features maps are upscaled into the HR residual image by Periodic Shuffling (PS) [2]. Finally, the HR output is estimated by adding the HR residual image to the bicubic interpolated LR image.

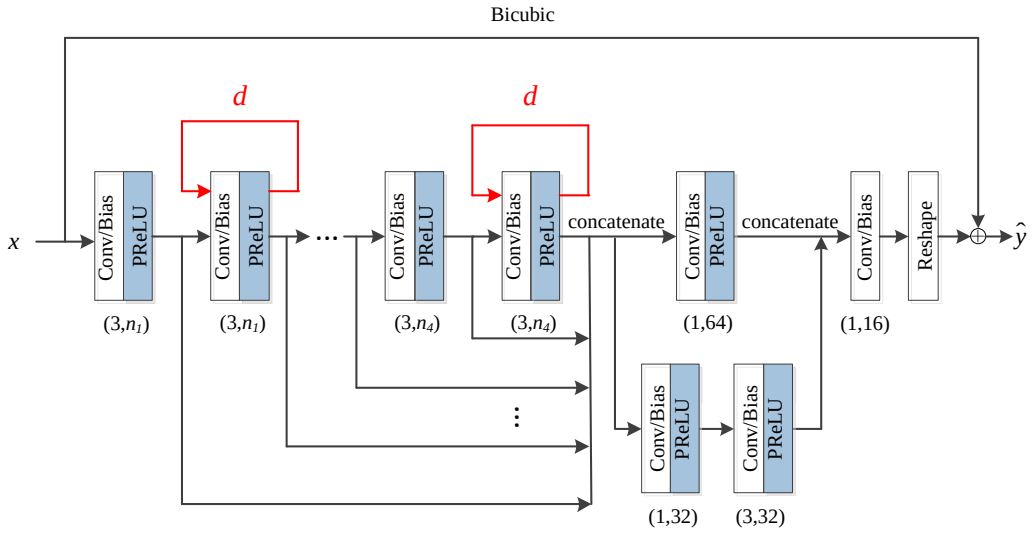

Fig. 2 Architecture of the proposed model when the scale factor is 4.

The red arrows refer to recursive convolutional layer and d is the number of recursions. The first number in parentheses refers to the filter size and the second number refers to the filter number.

2.2 Feature Extraction Network

The feature extraction network consists of four recursive modules. Each module contains a normal convolutional layer and a recursive convolutional layer [4], both of which have the same number of filters. PReLU is used as the activation function to solve the 'dying ReLU' problem [9]. Specifically, the proposed recursive module is formulated as

$$h_i = R_i^{(d)}\big(N_i(h_{i-1})\big), \tag{1}$$

where i = 1, 2, 3, 4, h_{i-1} and h_i are the input and the output of the i-th recursive module, N_i and R_i denote the functions of the i-th normal convolutional layer and the i-th recursive convolutional layer, respectively, and d is the number of recursions. Specially, when i = 1, we have

$$h_0 = x. \tag{2}$$

On the premise of certain parameters, the filter number of each recursive module is optimized for best performance, and the structure with decreasing number of filters is finally adopted (i.e., $n_1 > n_2 > n_3 > n_4$ in Fig. 2). That is why the proposed model is called the Piecewise-Recursive

Convolutional Network (PRCN). Compared with SR models that have the same number of filters in each convolutional layer [3, 4, 6, 7], the proposed PRCN model can better extract the local features of the input LR image and thus results in better performance.

The output of each convolution layer is passed to the next layer and simultaneously skipped to the reconstruction network. Accordingly, the output dimension N of the feature extraction network is formulated as

$$N = 2 \cdot (n_1 + n_2 + n_3 + n_4). \tag{3}$$

2.3 Image Reconstruction Network

The dimension of the extracted features is rather large. A huge amount of computation is required when such features are directly used to reconstruct the HR image. Therefore, parallelized 1×1 CNNs [8] are applied to reduce the input dimension and selectively retain high-order features at different levels. Compared with the plain structure where convolutional layers are cascaded directly, the parallel structure can enhance the learning ability and reduce the computational cost. Finally, the sub-pixel CNN [2] is used to transform the final features into the HR residual image. The HR prediction is obtained by adding the HR residual image to the interpolated LR image. As with typical residual learning networks, PRCN is designed to focus on learning residual output and thus significantly improves the learning performance.

3. Experiments

3.1 Datasets

We randomly select 10,000 images in the CelebA dataset [10] for training and another 1,000 images for testing. We make sure that people in the testing set do not appear in the training set. Center cropping is applied to the selected images to remove the unnecessary background. The size of cropped HR images is 128×128 pixels. The scale factor is set to 4 in all experiments. Accordingly, the size of input LR images is 32×32 pixels.

Data augmentation [11] is performed on the training images. Each training image is rotated by 90°, 180°, 270° and flipped horizontally to make 7 additional augmented versions. The total number of training set is 80,000 and 40 patches are used as a mini-batch. RGB images are converted to YCbCr images and only Y-channel is processed.

3.2 Training Setup

MSE-loss is adopted for training and the weight decay is set to 0.0001. The method proposed by He et al. [9] is used for weight initialization and all biases and PReLUs are initialized to 0. Dropout [12] is also applied with p = 0.8. Adam [13] with an initial learning rate = 0.002 is used for optimization. Learning rate is decreased by a factor of 2 if the loss does not decrease for 5 epochs. If learning rate is less than 2×10^{-5}, the procedure is terminated. Training roughly takes 7 hours using one GTX 1070 GPU.

3.3 Study of (n_1, n_2, n_3, n_4) and d

In this subsection, we explore various combinations of (n_1, n_2, n_3, n_4) and d to construct different networks, and find the values to achieve the best network performance. First, to clearly show how the parameters (n_1, n_2, n_3, n_4) affect our network, we fix the number of recursions d to 1 and change the number of filters. (64, 64, 64, 64) is taken as the reference value of (n_1, n_2, n_3, n_4).

Under the condition that the total parameters of the network remain unchanged, we record the Peak Signal-to-Noise Ratio (PSNR) under different values of (n_1, n_2, n_3, n_4). The results are shown in Table 1. Accordingly, we choose (96, 68, 49, 32) as the values of (n_1, n_2, n_3, n_4).

Table 1 Performance comparisons under different values of (n_1, n_2, n_3, n_4)

number	(n_1, n_2, n_3, n_4)	PSNR (dB)	number	(n_1, n_2, n_3, n_4)	PSNR (dB)
1	(64, 64, 64, 64)	30.937	7	(96, 64, 54, 32)	30.931
2	(80, 63, 63, 48)	30.942	8	(96, 68, 49, 32)	30.956
3	(80, 68, 57, 48)	30.944	9	(96, 72, 43, 32)	30.949
4	(80, 72, 52, 48)	30.950	10	(96, 79, 32, 32)	30.899
5	(80, 75, 48, 48)	30.952	11	(112, 55, 55, 16)	30.897
6	(96, 60, 60, 32)	30.923	12	(112, 63, 42, 16)	30.905

Next, we determine the best value of *d*. We use PSNR and computation complexity [14] as evaluation criteria to compare the reconstruction performance and computational efficiency under the different values of *d*. the results are shown in Table 2.

Table 2 Comparisons of different recursions on Peak-Signal-to-Noise-Ratio and complexity

d	1	2	3	4	5
PSNR (dB)	30.96	30.98	31.01	30.99	30.97
Complexity [k]	318.2	474.1	629.9	785.7	941.6

Compared with the normal convolution layer (i.e., $d = 1$), applying the recursive structure can significantly improve the performance of super-resolution reconstruction. Both performance and efficiency are considered and the number of recursions *d* is finally set to 3. Some examples of the proposed PRCN are illustrated in Fig. 3.

(a)Original face images (b)Bicubic images (c)super-resolution results (d)Ground truth

Fig. 3 Examples of the proposed method

3.4 Comparisons with State-of-the-Art Methods

We use PSNR and computation complexity to compare the proposed model with several representative SR networks (SRCNN [1], ESPCN [2], VDSR [3], DRCN [4], SRResNet [6]). The results are shown in Table 3. We can see the proposed PRCN has a state-of-the-art reconstruction performance. It significantly outperforms SRCNN and ESPCN by 0.72 and 0.44 dB, while the computation complexity is 17, 495, and 3 times smaller than VDSR, DRCN and SRResNet respectively.

Table 3 Comparisons with other super-resolution algorithms on PSNR and complexity

Methods	SRCNN	ESPCN	VDSR	DRCN	SRResNet	PRCN
PSNR (dB)	30.29	30.57	30.80	30.82	30.97	31.01
Complexity [k]	918.0	39.4	10674.2	312332.8	1877.7	629.9

4. Conclusions

In this paper, we propose a Piecewise-Recursive Convolutional Network (PRCN) for face image super-resolution. Our network takes the original low-resolution images as the input and thus reduces the computational cost. We also optimize the number of filters and recursions in the feature extraction network to achieve better performance and faster computation. Experimental results prove that PRCN is a concise and superior model for fast and accurate face image super-resolution.

References

[1] Dong C, Chen C L, He K, et al. Learning a Deep Convolutional Network for Image Super-Resolution [J]. 2014, 8692:184-199.

[2] Shi W, Caballero J, Huszar F, et al. Real-Time Single Image and Video Super-Resolution Using an Efficient Sub-Pixel Convolutional Neural Network[J]. 2016:1874-1883.

[3] Kim J, Lee J K, Lee K M. Accurate Image Super-Resolution Using Very Deep Convolutional Networks[C]// Computer Vision and Pattern Recognition. IEEE, 2016:1646-1654.

[4] Kim J, Lee J K, Lee K M. Deeply-Recursive Convolutional Network for Image Super-Resolution [J]. 2015:1637-1645.

[5] Dong C, Chen C L, Tang X. Accelerating the Super-Resolution Convolutional Neural Network [J]. 2016:391-407.

[6] Ledig C, Wang Z, Shi W, et al. Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network[J]. 2016:105-114.

[7] Mao X J, Shen C, Yang Y B. Image Restoration Using Convolutional Auto-encoders with Symmetric Skip Connections [J]. 2016.

[8] Szegedy C, Liu W, Jia Y, et al. going deeper with convolutions [J]. 2014:1-9.

[9] He K, Zhang X, Ren S, et al. Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification [J]. 2015:1026-1034.

[10] Liu Z, Luo P, Wang X, et al. Deep Learning Face Attributes in the Wild[C]// IEEE International Conference on Computer Vision. IEEE Computer Society, 2015:3730-3738.

[11] Timofte R, Rothe R, Gool L V. Seven Ways to Improve Example-Based Single Image Super Resolution [J]. 2015:1865-1873.
[12] Hinton G E, Srivastava N, Krizhevsky A, et al. Improving neural networks by preventing co-adaptation of feature detectors[J]. Computer Science, 2012, 3(4): 212-223.
[13] Kingma D, Ba J. Adam: A Method for Stochastic Optimization [J]. Computer Science, 2014.
[14] Jin Y, Kuwashima S, Kurita T. Fast and Accurate Image Super Resolution by Deep CNN with Skip Connection and Network in Network [J]. 2017:217-225.