



ISSN: 2377-6242 (Print)
ISSN: 2377-8156 (Online)

Applied Computer Letters (ACL)

DOI: <http://doi.org/10.7508/acl.01.2018.01.04>



ARTICLE

IMPROVEMENTS OF THE IMAGE SUPER-RESOLUTION ALGORITHM BASED ON CONVOLUTIONAL NEURAL NETWORKS

Claude M. Siefert*

College of Engineering, The University of Alabama, Tuscaloosa AL 35487, United Kingdom

*Corresponding Author E-mail: Siefert@hotmail.com

This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

ARTICLE DETAILS

Article History:

Received 3 February 2018
Accepted 2 March 2018
Available online 29 March 2018

ABSTRACT

Face hallucination is a branch of image super-resolution, which develops domain specific prior knowledge with strong cohesion to face domain. As the development of machine learning, there are numerous learning-based methods which have been proposed to solve the face hallucination problem. Learning based algorithms have been seen to achieve higher magnification factor with better visual quality than the other super resolution techniques such as bi-cubic interpolation and reconstruction based techniques. In the paper, we apply the deep learning theory to illusory face hallucination reconstruction. The model of deep convolution neural network is improved, the convolution neural network is added to the pool layer, the convolution kernel size is adjusted, the parameters are reduced, and the operation speed is increased. Finally, the iterative back projection method is used to reconstruct the face image after post-processing.

KEYWORDS

Image super-resolution, face hallucination, iterative back projection algorithm.

1. INTRODUCTION

Image super-resolution is a classical problem in the domain of computer vision. It aims to infer an HR image with crucial information from the given LR images. Face hallucination is a branch of image super-resolution, which develops domain specific prior knowledge with strong cohesion to face domain. It was first introduced by Baker and Kanada and has attracted growing attention due to practical importance in many face based applications such as face recognition, face alignment and so on. As the development of machine learning, there are numerous learning-based methods which have been proposed to solve the face hallucination problem. Learning based algorithms have been seen to achieve higher magnification factor with better visual quality than the other super resolution techniques such as bi-cubic interpolation and reconstruction based techniques [1, 2].

The algorithm based on Interpolation is the most basic method in face super-resolution research, including the nearest neighbor interpolation, bilinear interpolation, bicubic Interpolation etc. The method based on reconstruction has a fast speed and made a little improvement in image quality. However, because it is limited by the original information of the image, the ambiguity caused by low resolution sampling cannot be removed. Feerman et al. proposed a method of image super-resolution based on sample learning, and obtained the vector set in the external database through the nearest neighbor search. Timofte combined the sparse learning dictionary and the domain embedding method, and proposed the method of fixed domain regression, which improves the processing speed of the algorithm. Dong et al., proposed a deep convolutional neural network (CNN) based single image super

resolution method and showed that the traditional sparse coding-based algorithm can also be seen as a kind of deep convolutional network [3-5]. The end-to-end mapping between LR images and HR images was optimized in Dong's SR method, which achieves excellent reconstruction performance.

Inspired by the above literature, we apply the deep learning theory to illusory face hallucination reconstruction. The model of deep convolution neural network is improved, the convolution neural network is added to the pool layer, the convolution kernel size is adjusted, the parameters are reduced, and the operation speed is increased. Finally, the iterative back projection method is used to reconstruct the face image after post-processing [6-9].

2. METHODOLOGY

Image acquisition process may be affected by motion blur, optical blur, signal aliasing caused by down sampling and all kinds of noise. The picture is polluted by all of the above. Elad proposed a matrix vector approach to describe low resolution image imaging models.

$$y = HDXw + n \quad (1)$$

X represents high resolution images, y means low resolution images, and N represents additive Gauss noise. D, H and w denote the down sampling matrix, the fuzzy matrix and the geometric transformation matrix respectively. Hallucination face is the inverse process of face image degradation. The purpose is to give the low resolution image y to restore the original high resolution image X.

Convolutional Neural Networks are a biologically inspired variant of multi-layered perceptron networks (MLP's), specialized for image processing. First popularized by LeCun et al. in they are similar to other hierarchical feature extraction methods such as the Neocognitron and HMAX.

The structure of a typical CNN consists of alternating layers of convolutional and pooling followed by an output classification layer. Each type of layer contains several feature maps, or groups of neurons, in a rectangular configuration. The receptive field itself is simply a number of weighted connections, that is, each connecting edge has a weight. The group of weights applied by a neuron is called a weight kernel. A distinguishing property of these networks is that all neurons in a feature map share the same weight kernel. The idea behind this configuration is that a spatial feature detector should be useful across an entire image, instead of just at a particular location; for example a vertical edge detector. The convolutional layers in the network perform the majority of processing in these networks, with the feature maps in the pooling layers simply down-sampling their corresponding feature map in the convolutional layer [10].

In the past few years, the method based on deep learning has been improved and updated. It is not only applied to the image classification in the field of computer vision, but also from face recognition to semantic segmentation. Recently, deep learning method has also been applied to low level vision tasks, including image denoising, image enhancement, image super-resolution and so on. The seminal work of image super-resolution convolutional neural network (SRCNN) was done by Dong et al.

The model is mainly composed of three volumes, which generally simulate a sparse layer. Three coiling layers accomplish the following tasks: patch extraction and representation, nonlinear mapping and reconstruction.

1) Patch extraction and representation: A patch is extracted from a low resolution image and each patch is represented as a high dimensional vector. These high dimensional vectors are composed of a set of feature maps, and the dimension of the vector is equal to the number of maps.

2) Non-linear mapping: this step nonlinearly maps each high dimensional vector onto another high dimensional vector. Each mapped vector is conceptually the representation of a HR patch. These vectors comprise another set of feature maps.

3) Reconstruction: this step aggregates the above HR patch-wise representations to generate the final HR image. This image is expected to be similar to the ground truth R.

2.1 Iterative back projection algorithm

The iterative back projection algorithm proposed by Irani is the representative method of the original image restoration. It is used as the post-processing in the processing of the image super-resolution algorithm. The result of the single image interpolation is usually used as the initial solution of the high resolution image. According to the system model, the quasi low resolution image can be expressed as:

$$y_0 = HX_0 + n \quad (2)$$

If the x is equal to the original high resolution image and the upper analog imaging process conforms to the actual situation, the analog low resolution sequence y_0 is the same as the actual low resolution image y , and if it is different, the difference between y and y_0 is projected back to x_0 for correction.

In IBP algorithm, HR image is obtained by utilizing the backward projection of the error projection matrix based on the difference between simulated LR images and the observed LR images with up-sampling, reverse blur filter and reverse motion transform. This process iteratively calculates the estimated HR image until the energy of the error comes to a minimum value, or the iteration step achieves the maximum number. The simulated LR images, the error projection matrix for the acquisition process, and the IBP algorithm iterative reconstruction process can be written as:

$$E^{(n)} = \frac{1}{N} N \sum_{k=1}^N H \frac{BP}{k} \left(y_k - y^{\wedge \frac{(n)}{k}} \right) \quad (3)$$

$$z^{\wedge (n+1)} = z^{\wedge (n)} + \lambda E^{(n)} \quad (4)$$

where z^{\wedge} and $z^{\wedge (n+1)}$ denote the super resolution image gained from the (n) th and $(n+1)$ th iteration respectively, $y^{\wedge \frac{(n)}{k}}$ denotes the (n) th simulated LR images of $z^{\wedge (n)}$ under the imaging degradation model, $E^{(n)}$ is the difference between the simulated LR images and the observed LR images, $H \frac{BP}{k}$ is the (k) th back projection operation and λ is the iteration step.

2.2 The hybrid algorithm

The reconstruction algorithm of the iterative back-projection algorithm is not outstanding, but it can be combined with other super resolution methods to improve the performance. In this paper, the super-resolution algorithm based on convolution neural network is improved and combined with the iterative back-projection algorithm, a new composite algorithm is proposed.

2.2.1 The convolution layer

In the convolution layer, we mainly consider the influence of the size and number of convolution kernel on the processing effect and processing speed of the model, Convolution neural network proposed by Dong et al. And super-resolution model have three volumes of convolution kernel size of 9×9 , 1×1 , 5×5 . The improved algorithm in this paper has a three layer convolution kernel size of 3×3 , 1×1 , 3×3 . As shown in the following Figure 1.

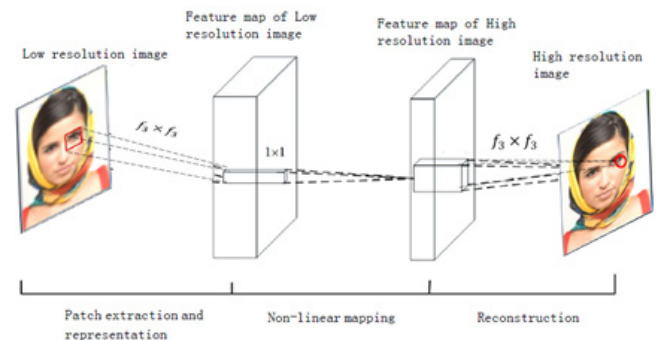


Figure 1: The flow chart of super-resolution algorithm.

The larger the convolution kernel size is in convolution, the better the super-resolution effect, but it will also increase the corresponding computation. In our improved convolutional neural network model, the first layer convolution kernel is changed to 3×3 , which can effectively reduce the number of parameters and cover the features of the image. If the second layer convolution kernel increases, the parameter will increase. The convolution kernel size of second layers is still 1×1 . The size of the convolution kernel of third layers is 3×3 . As the number of convolution cores may affect the quality of super-resolution images, we train and test the number of iterations of the convolution neural network model of 3-1-3 and 3-1-5 respectively. The experimental results show that each iteration 100 times 3-1-5 model takes 7.1 seconds more than the 3-1-3 model; The average PSNR 3-1-5 model of the phase output super resolution image is 0.1dB higher than that of the 3-1-3 model. Considering the processing results and computing speed, we select the 3-1-3 model.

The number of convolution kernel and convolution kernel size determines the super resolution effect together. In the super-resolution algorithm 9-1-5 model proposed by Dong and others, the first convolution kernel number and the second layer convolution kernel number im-

provement algorithm are tested in the 3-1-3 model for different size and the final selection $n_1 = 64, n_2 = 64$.

According to the selection of the number of convolution kernel we chose and the test image under the condition of the same iteration number and the same learning rate in the improved algorithm 3-1-3 model. The PSNR value of super resolution image test is 5×10^4 and the learning rate is 10^{-3} . The comparison data obtained are shown in Table 1.

Table 1: The test PSNR values of different convolution kernel numbers.

Kernel number	$n_1 = 64, n_2 = 32$	$n_1 = 64, n_2 = 64$	$n_1 = 128, n_2 = 64$
PSNR	33.57	33.65	30.14

At the same time, the influence of the amount of calculation and information on the speed of the image super-resolution is considered. The size of the patch is selected to be 33×33 , and the size of the sub image is properly increased. Experimental results show that the increase of input block size improves training speed and shortens training time.

2.2.2 Pooled layer

Dong et al. proposed to apply convolution neural network model to image super-resolution processing, there is no pool layer in the model. In addition to improving the size and number of convolution kernels, we also introduce pooled layers. After introducing the pool layer into the first, second volume layer, the introduction of the pool layer can reduce the output vector, reduce the dimension and speed up the training. The phenomenon of over fitting can be avoided after adding the pool layer. After introducing the pool layer, the network depth reaches five layers, and the network structure with deeper layers is more conducive to the learning of image data. At the same time, the number of parameters can be reduced by pooling layer.

To make the pooling unit have translation invariance, that is, after a small translation, the image still produces the same characteristics as before. We can choose the continuous range in the image as the pool area and only pool the features generated by the same hidden unit, so the number of the pool layer input feature graph will not change, but the size of the feature graph will be reduced. This process is actually a down sampling process. It can be expressed as a formula:

$$X_j^k = f \left[\beta_j^k \text{down} \left(x_j^{k-1} + b_j^k \right) \right] \quad (5)$$

2.3 Architecture of hybrid algorithm

The whole model of the hybrid algorithm is divided into nine layers, including four coiling layers, two pool layers, two subsampling layers (one for lower sampling, the other for upper sampling), and the other for differential stratification. The detailed description of the specific functions of each layer is as follows:

The first five layers are the framework of the super-resolution algorithm of the convolution neural network (SRCNN model), which are mainly implemented in the five layers: the extraction and expression of the patch, the nonlinear mapping and reconstruction.

Down-sampling layer. This operation down-samples the image derived from the third layer. As a result, a LR version of reconstruction image is obtained.

Difference layer. This operation calculates the difference between down-sampling version of the original HR image and the corresponding

counterpart we acquired above. The difference is treated as reconstruction simulation error, and it also can be considered as a prior guidance that has been introduced Up-sampling layer. This operation up-samples the simulation error to generate the simulation error of HR version.

Update layer. This operation performs a convolution with the above simulation error, and then the final HR image is updated based on the synthesis of the third layer's result with the convolution version of simulation error.

3. EXPERIMENTAL RESULTS

Considering the test and contrast, the image degradation model we adopted is to generate low resolution images by sampling the original high-resolution images after Gauss blur. The initial high resolution image is input into the convolution neural network model of off-line training, and the final high resolution image is generated by the processing of network learning and optimization after the processing of the model parameters. In order to ensure the objectivity and rationality of the comparison, the learning results of the network model with the same iteration number are selected for comparative test. The maximum iteration number of the composite algorithm is 5×10^5 due to device constraints.

We choose a Bicubic method, the NE algorithm, the ScSR algorithm and the SRCNN algorithm for comparative test. We evaluate and compare the performance of our proposed models in terms of peak signal to noise ratio (PSNR) and structure similarity (SSIM). The model is tested on 500 images of BANVA dataset. Here are the visual qualities of the face hallucination results generated by our method and other competing ones in Figure 2.



Figure 2: Results of images with different algorithms.

Table 2 shows the comparison results in terms of PSNR and SSIM between the proposed improved method and some competing methods. The best results are high-lighted in bold. As one can see, our proposed method improves the results both in terms of PSNR and SSIM. And they all have an up-scaling factor of in our experiments. We can make some observations from the results. Obviously, the proposed modified network has already achieved better results than the competing methods, which validates that introducing image prior for face hallucination works well.

Table 2: The comparison results in terms of PSNR and SSIM between the proposed improved method and some competing methods.

Image	BI		NE		ScSR		SRCNN		Proposed	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
1	27.50	0.798	28.30	0.803	29.27	0.825	30.23	0.834	31.01	0.871
2	28.91	0.837	28.36	0.844	29.89	0.874	30.59	0.882	31.87	0.912
3	28.01	0.823	28.81	0.832	30.61	0.853	31.02	0.875	32.23	0.891
4	27.04	0.812	28.36	0.829	28.69	0.843	30.22	0.862	31.35	0.886
5	28.70	0.833	29.09	0.838	29.21	0.853	31.21	0.880	31.93	0.879
6	32.48	0.897	32.56	0.898	33.45	0.905	34.00	0.914	34.08	0.917
Average	28.77	0.833	29.25	0.840	30.197	0.859	31.21	0.875	32.08	0.890

4. CONCLUSION

In this paper, by analyzing the training process of convolutional neural networks, we have made a series of improvements to the image super-resolution algorithm based on convolutional neural networks. Compared with the traditional algorithm, the results show that the improved algorithm has better reconstruction effect, higher edge sharpness and clearer picture. Improved convolutional neural network algorithm can achieve better results with less iteration and significantly reduce training time.

REFERENCES

- [1] Uno, S., Kamiya, M., Morozumi, A., Urano, Y. A green-light-emitting, spontaneously blinking fluorophore based on intramolecular spirocyclization for dual-colour super-resolution imaging. *Chemical communications (Cambridge, England)* (2017) 54(1): 102-105.
- [2] Lee, J., Park, S., Kang, W., Hohng, S. Accelerated super-resolution imaging with fret-paint. *Molecular Brain* (2017) 10(63).
- [3] Obuchi, T., Ikeda, S., Akiyama, K., Kabashima, Y. Accelerating cross-validation with total variation and its application to super-resolution imaging. *Plos One* (2017) 12(e018801212).
- [4] Yang, W., Feng, J., Yang, J., Zhao, F., Liu, J., Guo, Z., Yan, S. Deep Edge Guided Recurrent Residual Learning for Image Super-Resolution. *IEEE Transactions On Image Processing* (2017) 26(12): 5895-5907.
- [5] Peedikakkal, L.V., Steventon, V., Furley, A., Cadby, A.J. Development of targeted STORM for super resolution imaging of biological samples using digital micro-mirror device. *Optics Communications* (2017) 404(SI): 18-22.
- [6] Huh, H., Lee, J., Kim, H.J., Hohng, S., Kim, S.K. Morphological analysis of oligomeric vs. fibrillar forms of alpha-synuclein aggregates with super-resolution BALM imaging. *Chemical Physics Letters* (2017) 690: 62-67.
- [7] Schueder, F., Lara-Gutierrez, J., Beliveau, B.J., Saka, S.K., Sasaki, H.M., Woehrstein, J.B., Strauss, M.T., Grabmayr, H., Yin, P., Jungmann, R. Multiplexed 3D super-resolution imaging of whole cells using spinning disk confocal microscopy and DNA-paint. *Nature Communications* (2017) 8(2090).
- [8] Lu, W., Sun, H., Wang, R., He, L., Jou, M., Syu, S., Li, J. Single image super resolution based on sparse domain selection. *Neurocomputing* (2017) 269(SI): 180-187.
- [9] Ma, Q., Hu, H., Huang, E., Liu, Z. Super-resolution imaging by metamaterial-based compressive spatial-to-spectral transformation. *Nanoscale* (2017) 9(46): 18268-18274.
- [10] Bi, C., Li, Y., Zhang, Y., Xu, L. Super-resolution imaging of low-frequency sound sources using a corrected monopole time reversal method. *Journal of Sound and Vibration* (2017) 410: 303-317.

