EasyChair Preprint

№ 694

Image Super-Resolution Reconstruction Method Based on Global and Local Residual Learning

Jingru Hou, Yujuan Si and Liangliang Li

December 23, 2018

Image Super-Resolution Reconstruction Method Based on Global and Local Residual Learning

Jingru Hou
College of Communication Engineering,
Jilin University,
Changchun,130012,China
2665541682@qq.com

Yujuan Si*
College of Communication Engineering,
Jilin University,
Changchun,130012,China
Zhuhai College of Jilin University,
Zhuhai,519041,China
siyj@jlu.edu.cn

Liangliang Li
College of Communication Engineering,
Jilin University,
Changchun,130012,China
leeliangliang@163.com

ABSTRACT

In the image super-resolution (SR) algorithm, some more effective methods are obtained by convolutional neural networks (CNN). However, at present, there are two main problems in the SR algorithm using CNN: First, due to the degradation of the image, it is easy to cause partial loss of image details in a very deep network. Second, although the depth network is very powerful, when performing efficient nonlinear mapping from low-resolution (LR) input images to high-resolution (HR) target images, a large number of parameters are required, which easily causes learning difficulties. Therefore, in this paper, combined global residual learning (GRL) and local residual learning (LRL)， we propose a new method to effectively obtain image details . Especially, using the stacked local residual block (LRB) structure for nonlinear mapping the parameters of CNN can be reduced and the image degradation problem can be overcome effectively. Due to the high correlation between LR images input and HR images output in the network, we can use the skip-connection method to reconstruct of HR images from most of the LR information. The experimental results show that the proposed image SR algorithm is effective.

CCS Concepts

• Computing methodologies → Image processing.

Keywords

super-resolution; convolutional neural network; global and local residual learning; local residual block; skip-connection

1. INTRODUCTION

Image super-resolution (SR) is the process of reconstructing a high-resolution (HR) image with one or more low-resolution (LR) images by image processing or signal processing. SR reconstruction can be widely used in various fields [1] [2], because HR images can provide clearer image details, contain more information, and have higher pixel density. Therefore, the image super-resolution reconstruction algorithm has attracted extensive attention and has strong research value.

At present, there are many methods for image SR. For example, there are some image super-resolution reconstruction algorithms based on interpolation. Classic interpolation algorithms include nearest neighbor interpolation, linear interpolation, bicubic interpolation, and spline interpolation [3]. These methods do not fundamentally increase the amount of information contained in the images. In the dictionary-based SR algorithm, the typical algorithm is sparse coding [4]. The learning-based image SR algorithm learns the similarity between high- and low-resolution image blocks by using a deep convolutional neural network (CNN). Therefore, the low-resolution image blocks to be processed can be more accurately generated with the corresponding high-resolution image blocks, which overcome the problem that it is difficult to learn accurate high-low resolution image blocks spatial mapping relationship by using the hand-designed features in the traditional sample-based SR algorithm. Dong C. et al. applied CNN to super-resolution reconstruction of images for the first time. The algorithm is called super-resolution by convolutional neural network (SRCNN) [5]. The network input is a LR image, and the output is a HR image, which realizes the end-to-end reconstruction of the image. All of the feature extraction and aggregation are done by the hidden layers of the CNN, but the SRCNN does not consider any self-similar properties. After the SRCNN model was proposed, more models for super-resolution neural networks were proposed. For example, the fast super-resolution convolutional neural networks (FSRCNN) is also proposed by Dong C. et al [6]. This method can directly process small-resolution images. The advantage is that reducing the pictures can shorten the training time. Subsequently Kim J. proposed deeply-recursive convolutional network for image super-resolution (DRCN) [9], which is divided into three layers. At the layer of nonlinear mapping, it uses a recursive network. After that, Tai Y. et al proposed the image super-resolution via deep recursive residual network (DRRN) [10]. It uses a deeper network structure to improve performance.

This paper mainly analyzes the image super-resolution reconstruction based on global and local residual learning (GLRL), aiming to improve the resolution of the image. By increasing the network width, a better reconstruction effect can be obtained, which lays a solid theoretical foundation for the application of image SR. In order to demonstrate the effectiveness of the proposed method, Fig. 1 shows the value of the peak signal to noise ratio (PSNR) of some image super-resolution algorithms. The larger the PSNR, the better the reconstruction effect. As the figure shows, the proposed method has a higher PSNR value than Bicubic, single image super-resolution from transformed self-exemplars (SelfEx), image super-resolution by mid-frequency sparse representation and total variation regularization (MFTV)

and SRCNN by 1.95, 0.06, 0.65, 0.28, respectively [12] [13]. The algorithm of this paper has achieved good results.

This paper is divided into five sections. The first section is the introduction, which introduces several existing methods of image SR. The second section introduces the network related to our work, the third section describes the details of the algorithm, the fourth section shows the experimental results, and the last section summarizes.

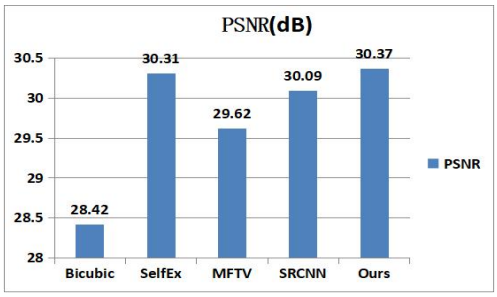

Figure 1. The PSNR of recent models for scale factor ×4 on Set5 .

2. RELATED WORK

This section highlights four networks related to our work: SRCNN, deep residual learning for image recognition (ResNet)[14], DRCN, DRRN.

2.1 SRCNN

SRCNN[5] is the first method to apply CNN to image SR. Its structure is divided into three parts.

The first step is the extraction and feature representation of the image patches, and the features of the image patches are extracted using the properties of the convolutional network. The formula is as follows:

$$F_1(Y) = \max(0, W_1 * Y + B_1) \tag{1}$$

The second step is a nonlinear mapping. The n1 dimensional feature of the first step is mapped to the n2 dimension. The formula is described as follows:

$$F_2(Y) = \max(0, W_2 * F_1(Y) + B_2) \tag{2}$$

The third step is image reconstruction. Reconstruction can be done by simply doing another convolution. The formula is computed by as follows:

$$F(Y) = W_3 * F_2(Y) + B_3 \tag{3}$$

2.2 ResNet

In this paper, the concept of residual learning is proposed for the first time, and this method is used to solve the problem of image degradation[10].

Suppose $H(x)$ is the desired underlying map, and x is the input to these layers, letting the stacked nonlinear layers fit another map:

$$F(x) = H(x) - x \tag{4}$$

The residual network structure is as follows:

$$y = \sigma(F(x, W) + x) \tag{5}$$

where x and y represent the input and output of the layer, respectively. The function $F(x, W)$ represents the learned residual mapping. $F = W_2\sigma(W_1 x)$, where σ represents the activation function ReLU[15][16]. By stacking these structures, a very deep 152-layer network was built, which was the first in the ILSVRC 2015 classification competition.

2.3 DRCN

DRCN[9] [10] uses a recursive convolution network. The algorithm is mainly divided into three parts, the first part is embedding network. The output $f_1(x)$ of this layer represents the feature map of a given image x, which is equivalent to feature extraction in SRCNN. The second is the inference network. The output $f_2(f_1(x))$ of this layer indicates that T recursion ($T = 16$) is superimposed in the recursive layer, and the weight is shared in the recursive layer, which is equivalent to the nonlinear transformation of features. The third is the reconstruction network, whose output is $f_{\text{Rec}}(H_T)$, where H_T is the output of the inference network. This network result can be expressed as:

$$y_t = f_{\text{Rec}}(f_2^{(t)}(f_1(x))) + x \tag{6}$$

where x is the skip connection, basically a global residual learning (GRL). Each intermediate predictive y_t is under supervision. Finally, using the integration strategy, the output is the weighted average of all predictions $y = \sum_{t=1}^{T} w_t \bullet y_t$, the weight of which is learned during the training process.

2.4 DRRN

DRRN[10] uses GRL in the identity branch, and introduces recursive learning into the residual branch by constructing a recursive block structure. Multiple residual units are superimposed in the structure using a multipath structure. All residual units share the same input for the identification branches. The residual unit is defined as:

$$H_u = G(H_{u-1}) = F(H_{u-1}, W) + H_0 \tag{7}$$

where G is the function of the residual unit and H_0 is the result of the first convolutional layer in the recursive block. Since the residual unit is recursively learned, the weight set W is shared between residual units within the recursive block, but differs between different recursive blocks. The superposition of a plurality of residual units constitutes a residual block, and the b-th residual block is expressed as:

$$x_b = H_b^U = R_b(x_{b-1}) = G^{(U)}(f_b(x_{b-1})) \tag{8}$$

where u-folding operation of G_b is performed. R is the function of the b-th recursive block.

The result of DRRN is expressed as:

$$y = D(x) = f_{\text{Rec}}(R_B(R_{B-1}(...(R_1(x))...))) + x \tag{9}$$

where B is the number of recursive block, f_{Rec} is a function of the last convolutional layer reconstruction residual in DRRN.

In Fig. 2, (a) is expressed as the structure of SRCNN; (b) is the simplified structure of Resnet. The picture shows two residual unit structures; (c) represents the structure of DRCN. The yellow dashed box refers to a recursive layer, among which the convolutional layers (with gray color) share the same weights; (d) represents the simplified structure of DRRN. The red dashed box refers to a recursive block consisting of two residual units. In the recursive block, the corresponding convolutional layers in the residual units (with green or yellow color) share the same weights.

$\oplus$ is the element-wise addition.

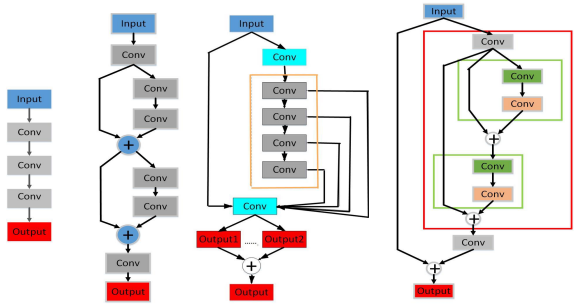

Figure 2. (a) SRCNN. (b) ResNet. (c) DRCN. (d) DRRN.

3. METHOD

In this paper, we propose a new algorithm called image super-resolution reconstruction based on global and local residual learning (GLRL). In the image reconstruction, the three basic models mentioned by SRCNN are used, and the GRL and LRL methods are combined. The network structure is inspired by the above four algorithms. It is shown in Fig. 3. Next, we elaborate on this algorithm.

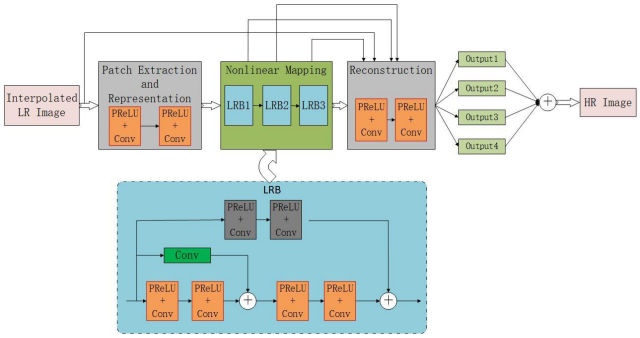

Figure 3. The proposed super-resolution image reconstruction framework.

3.1 Patch Extraction and Representation

After bicubic interpolation and amplification of the input LR image, the feature representation of the image patch is performed by using a two-layer convolution network. After each convolutional layer, we use the parameter rectified linear unit (PReLU) as the activation function instead of the commonly used rectified linear unit (ReLU) [17]. Because PReLU can avoid the "death feature" caused by the zero gradient in the ReLU. PReLU is defined as:

$$\sigma(x_i) = \max(x_i, 0) + a_i \min(x_i, 0) \tag{10}$$

where x_i is the input signal of the i th convolutional layer and a_i is the coefficient of the negative part. It is worth noting that a_i is based on data for PReLU and zero for ReLU. Apply the pre-activation strategy to the network, which was proposed by He et al [18].

$$x_i = F_i(\sigma(x_{i-1}), W_i) \tag{11}$$

where x_{i-1} and x_i are the input and output of the i th layer network, f_i is the mapping function, and w_i is a set of weights.

3.2 Nonlinear Mapping

Nonlinear mapping is the core step of this algorithm. We implement nonlinear mapping by stacking local residual blocks (LRB) [17]. The superposition of LRB can partially overcome the degradation problem of the image and reduce network redundancy by increasing the network width.

Fig. 4 shows a comparison of the partial structure of DRRN and the LRB. The DRRN structure in this Fig. 4(a) is a structure diagram in which the number of residual units in the recursive block is two. As is shown in Fig. 4(b), the construction of the LRB is based on DRRN structure, and the shallow convolution branch and the deep convolution branch are added. Both deep convolution branches and shallow convolution branches use LRL, and shallow branches can provide more high frequency features for deep branches. LRB can be written as:

$$x^u = U(x^{u-1}) = G_D(\sigma(x^{u-1}), \theta_D) + G_S(\sigma(x^{u-1}), \theta_S) \tag{12}$$

where U represents the function of the residual unit, x^{u-1} and x^u represent the input and output, respectively, and θ_D and θ_S represent the weight sets of the deep and shallow branches, respectively. G_D and G_S are residual mappings that need to be learned for deep and shallow branches, respectively.

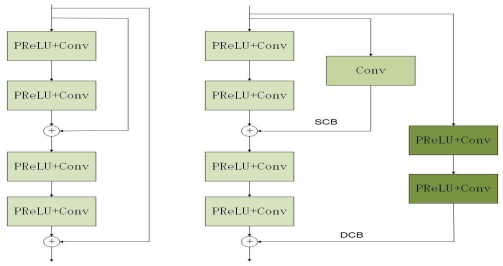

Figure 4. (a) the partially simplified structure of DRRN; (b) the structure of LRB with a shallow convolutional branch (SCB) and a deep convolutional branch (DCB).

3.3 Image Reconstruction

The layer of SR reconstruction is accomplished by a two-layer convolution network, which structure is similar with the image patch extraction and presentation module. But the input of this layer is the interpolated LR image and the output of each LRB, and the weighted average of each output after the reconstructed layer is obtained to achieve our final HR image.

In order to optimize the proposed network parameters, we need to minimize the loss between the predicted HR image and the ground truth image. The loss function is expressed as:

$$L(\Theta) = \frac{1}{2N} \sum_{i=1}^{N} \left\| \tilde{x}^{(i)} - D(x^{(i)}) \right\|^2 \tag{13}$$

Where N is the number of training patches, Θ is the parameter set, $\tilde{x}^{(i)}$ is the ground truth image, and D is the mapping function of the HR image generated from the interpolation low resolution image training.

4. EXPERIMENTS AND RESULTS

The image super-resolution reconstruction based on global and local resident learning (GLRL) proposed in this paper uses 291 pictures as the data set for training. 91 pictures are from the training set adopted by Yang et al [20], the other 200 pictures are

from Berkeley Segmentation data set [21]. And the Set5 [22], Set14 [23], and BSD100 [24] are used as test sets, containing 5, 14 and 100 images, respectively. These images are processed through our improved model to produce the final HR image.

4.1 Experimental Details

This experiment was performed on 291 training datasets and trained using scale enhancements with scales of ×2, ×3, and ×4, respectively. We provide parameters for training the final model using a network depth of 16 layers, a mini-batch size of 32, a momentum and weight decay parameter set to 0.9 and 0.0001, respectively. The initial learning rate is set to 0.1. There are 76 epochs (6536 iterations) in all the experiments we trained. Then every 10 epochs learning rates are cut in half. This experiment was conducted in a Caffe environment, and the GPU used 1080Ti.

4.2 Discussion on the Number of LRB

In this section, we analyze the effect of the number of LRB [17] on the image reconstruction effect. The experiment compares the effect of two LRB and three LRB on the Set5. The results are shown in Tab. 1. The average PSNR and SSIM of the three LRB are much higher than the values of the two LRB. It can be seen that the more the number of LRB, the better the reconstruction effect of the image, if the GPU allows it. In other words, increasing the depth and width of the network at the same time helps to restore image detail. The effect of the reconstructed image of 3LRB is significantly better than that of 2LRB.

Table 1. Comparison of PSNR and SSIM between the two LRB and three LRB

Set5	2LRB	3LRB
	30.36/0.8746	33.09/0.9124

4.3 Comparison with Other Methods

We compare our method to the other four image SR methods, which are named Bicubic [10], SelfEx [12], MFTV [13], SRCNN [5]. The implementation of these methods is based on their source code. In order to make a fair comparison of the quality of recovery, all models are trained on 291 image data sets. Tab. 2 lists the quantitative results on the test set Set5, Set14 and BSD100 under different scale factors (×2, ×3, ×4). The objective evaluation criteria usually have two: peak signal to noise ratio (PSNR), structural similarity index (SSIM) [10]. The larger the two values, the better the reconstruction effect. As seen from Tab. 2, the values of PSNR and SSIM of our proposed method are significantly higher than those of other methods. It can be seen that the reconstruction effect of our method is better than the other four methods. As can be seen from Fig. 6, in our approach, the outlines are clean and vivid, closest to the ground truth image, while in other methods they are heavily blurred or distorted.

Figure 6. Qualitative comparison. The first row shows image "baby" (Set5 with scale factor ×3). The second row shows image "butterfly" (Set5 with scale factor ×3).

5. CONCLUSIONS

In this experiment, we propose an image super-resolution reconstruction based on GLRL. In this network, we use a skip-connection to use LR images as input to the reconstruction layer for GRL, and perform non-linear mapping by superposition of LRB [17], and learn the output of the LRB as an input of the reconstruction layer. This helps to better restore image details. Compared with other methods, the experimental indicators have certain advantages in PSNR and SSIM, so our method can obtain better reconstruction effect.

Table 2. Comparison of PSNR and SSIM values for various methods with a scale of ×2,×3,×4 on Set5, Set14, BSD100

Dataset	Scale	Bicubic	SelfEx [12]	MFTV [13]	SRCNN [5]	Proposed
Set5	×2	33.66/0.9299	36.49/0.9537	——	36.34/0.9542	36.54/0.9550
	×3	30.39/0.8682	32.58/0.9093	32.12/0.9032	32.39/0.9090	33.09/0.9124
	×4	28.42/0.8104	30.31/0.8619	29.62/0.8561	30.09/0.8628	30.37/0.8635
Set14	×2	30.24/0.8688	32.22/0.9034	——	32.45/0.9067	32.56/0.9102
	×3	27.55/0.7742	29.16/0.8196	29.05/0.8653	29.00/0.8215	29.59/0.8705
	×4	26.00/0.7027	27.40/0.7518	27.07/0.8000	27.50/0.7513	27.53/0.8102
BSD100	×2	29.56/0.8431	31.18/0.8855	——	31.36/0.8879	31.41/0.8890
	×3	27.21/0.7385	28.29/0.7840	27.99/0.7794	28.41/0.7863	28.59/0.7891
	×4	25.96/0.6675	26.84/0.7106	26.12/0.7152	26.90/0.7101	26.97/0.7190

Comment: Red color indicates the best performance of our methods and blue color indicates the best performance of previous method

6. ACKNOWLEDGMENTS

This work was supported by the Key Scientific and Technological Research Project of Jilin Province under Grant Nos. 20150204039GX and 20170414017GH; the Natural Science Foundation of Guangdong Province under Grant No. 2016A030313658; the Innovation and Strengthening School Project (provincial key platform and major scientific research project) supported by Guangdong Government under Grant No. 2015KTSCX175; the Premier-Discipline Enhancement Scheme Supported by Zhuhai Government under Grant No. 2015YXXK02-2; the Premier Key-Discipline Enhancement Scheme Supported by Guangdong Government Funds under Grant No. 2016GDYSZDXK036.

7. REFERENCES

[1] M. W. Thornton, P. M. Atkinson, D. A. Holland. Sub-pixel mapping of rural land cover objects from fine spatial resolution satellite sensor imagery using super-resolution pixel-swapping. International Journal of Remote Sensing, 2006, 27(3):473-491.

[2] Shi W, Caballero J, Ledig C, et al. Cardiac image super-resolution with global correspondence using multi-atlas patch match. Med Image Comput Comput Assist Interv, 2013, 16(3):9-16.

[3] Temizel, Vlachos. Wavelet domain image resolution enhancement. IEE Proceedings - Vision Image and Signal Processing, 2006, 153(1):25-30.

[4] Yang J, Wright J, Huang T S, et al. Image super-resolution via sparse representation. IEEE Trans Image Process, 2010, 19(11):2861-2873.

[5] Dong C, Chen C L, He K, et al. Learning a deep convolutional network for image super-resolution. Computer Vision,ECCV 2014 - 13th European Conference, 2014, 8692:184-199.

[6] Dong C, Chen C L, Tang X. Accelerating the super-resolution convolutional neural network. Computer Vision - 14th European Conference ,2016:391-407.

[7] Shi W, Caballero J, Huszar F, et al. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In Proceedings of 2016 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA ,27-30,JUN,2016:1874-1883.

[8] Kim J, Lee J K, Lee K M. Accurate image super-resolution using very deep convolutional networks. In Proceedings of 2016 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA ,27-30,JUN,2016:1646-1654.

[9] Kim J, Lee J K, Lee K M. Deeply-recursive convolutional network for image super-resolution. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR),Seattle, WA ,27-30,JUN,2016:1637-1645.

[10] Tai Y, Yang J, Liu X. Image super-resolution via deep recursive residual network. In Proceedings of 30th IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR),Honolulu, HI, 21-26,July,2017:2790-2798.

[11] Tong T, Li G, Liu X, et al. Image super-resolution using dense skip connections. In Proceedings of 16th IEEE International Conference on Computer Vision (ICCV),Venice, ITALY, 22-29, OCT,2017:4809-4817.

[12] Huang J B, Singh A, Ahuja N. Single image super-resolution from transformed self-exemplars. In Proceedings of 2015 IEEE Conference on Compute r Vision and Pattern Recognition (CVPR), Boston MA, 07-12, June, 2015, 5197-5206.

[13] Jian X, Chang Z, Fan J, et al. Image superresolution by midfrequency sparse representation and total variation regularization. Journal of Electronic Imaging, 2015, 24(ISFA):013039.

[14] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition. In Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA ,27-30,June,2016:770-778.

[15] Matthew D. Zeiler, Rob Fergus. Visualizing and understanding convolutional networks. In Proceedings of 13th European Conference on Computer Vision (ECCV),Zurich,SWITZERLAND,06-12,SEP,2013, 8689:818-833.

[16] Dong C, Chen C L, He K, et al. Image super-resolution using deep convolutional networks. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2016, 38(2):295-307.

[17] Shi J, Liu Q, Wang C, et al. Super-resolution reconstruction of MR image with a novel residual learning network algorithm.. Physics in Medicine & Biology, 2018,63(8):085011.

[18] He K, Zhang X, Ren S, et al. Identity mappings in deep residual networks. In Proceedings of 14th European Conference on Computer Vision (ECCV),Amsterdam, NETHERLANDS,08-16,OCT,2016:630-645.

[19] Zagoruyko, Sergey, Komodakis, Nikos. Wide residual networks. In Proceedings of British Machine Vision Conference (BMVC),2016: 87.1-87.12 .

[20] Yang J, Wright J, Huang T, et al. Image super-resolution as sparse representation of raw image patches. In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition,Anchorage,AK, 23-28, JUN,2008:2378-+

[21] Martin D, Fowlkes C, Tal D, et al. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In Proceedings of the IEEE International Conference on Computer Vision, 2001:416-423 vol.2.

[22] Bevilacqua C M , Roumy A , and Morel M-L A. Lowcomplexity single-image super-resolution based on non-negative neighbor embedding. In Proceedings of British Machine Vision Conference (BMVC), Oct,2012:216

[23] Roman Zeyde, Michael Elad, Matan Protter. On single image scale-up using sparse-representations. International Conference on Curves and Surfaces. Springer-Verlag, 2012:711-730.

[24] Martin D, Fowlkes C, Tal D, et al. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In Proceedings of the IEEE International Conference on Computer Vision, 2001:416-423 vol.

Authors' background

Your Name	Title*	Research Field	Personal website
Jingru Hou	master student	Image processing	2665541682@qq.com
Yujuan si	full professor	Signal processing and pattern recognition	siyj@jlu.edu.cn
Liangliang Li	Phd candidate	Image processing	leeliangliang@163.com

*This form helps us to understand your paper better, the form itself will not be published.

*Title can be chosen from: master student, Phd candidate, assistant professor, lecture, senior lecture, associate professor, full professor