An improved denoising model for convolutional neural network

Wenjing Wang
School of Information and Communication Engineering Communication University of China, Beijing, China
wwj13269133866@163.com

Abstract. In recent years, computer vision and many other research fields have put forward higher and higher requirements for image quality, and image denoising has become an important research direction. Aiming at the problems of general image denoising effect and long algorithm convergence time, this paper makes an improvement on the original network structure of denoising convolutional neural network DnCNN model and adds channel attention mechanism in the penultimate layer. The channel attention mechanism further improves the signal-to-noise ratio (SNR) and optimizes the noise reduction effect through convolution and deconvolution neural networks. The experimental results show that, under the same noise reduction level of 25, the PSNR of the proposed DnCNN2 algorithm is 1.25dB and 1.62dB higher than that of the DnCNN algorithm after the verification of the BSD68 and SET12 test data sets. Structural similarity (SSLM) was increased by 0.0011 and 0.0103, respectively. Meanwhile, the convergence time of the algorithm becomes shorter. Compared with traditional deep learning image denoising algorithms, the noise reduction effect and operation efficiency of the proposed method are also improved, which has certain advantages.

Keywords: image denoising; Deep convolutional neural networks; Residual learning; Attentional mechanism.

1. Introduction
Due to the internal factors of the image acquisition system, the environmental impact of the transmission process, and human interference, it is extremely difficult to completely avoid the generation of image noise, which will lead to the deterioration of the quality of the image acquisition, and affect the subsequent processing process such as feature extraction and image segmentation. Therefore, image denoising is of great significance.

For the past few years, in recent years, with the development of the computer vision field, image denoising methods based on deep learning have been developed continuously. Relying on the superior data fitting ability and exquisite structure design of the artificial neural network, image denoising technology based on convolutional neural network breaks through the constraints of some traditional image denoising methods and shows the excellent effect of noise reduction.
Zhang et al. [1] proposed feed-forward denoising neural networks (DnCNNs) based on deep convolution by combining batch normalization and residual learning techniques, and the denoising effect was remarkable. Through a feed-forward convolutional neural network, DnCNN separates the noise from the noise image, and through the mutual integration of residual learning and batch normalization, improves the denoising performance while speeding up and stabilizing the training process. Experiments also confirm that DnCNN’s denoising effect is better than many traditional denoising algorithms.

However, in the process of convolutional neural network training, with the increase of the number of iterations, the convergence performance begins to gradually decrease, and there is still a certain gap between the generated denoising image and the clean image. Given these two aspects of the shortcomings of research innovation.

The DnCNN2 algorithm proposed in this paper expands the advantages of DnCNN and enhances the noise reduction effect of the model. This algorithm will add a channel attention mechanism before the penultimate layer of the DnCNN model so that the information flow between modules is more sufficient and the feature learning is more perfect. At the same time, the model has a relatively shallow depth (18 layers), which is suitable for GPU parallel computing and has higher computational efficiency.

2. Related Work

2.1. The noise reduction task

The purpose of image denoising is to reduce the pollution of noise to the image as far as possible. Through the reconstruction of the image information itself, the clean image that is close to or the same as the original image can be recovered. However, due to the diversity of noise in nature and the abundance of statistical models of noise, real noise images, and the corresponding clean image data sets are scarce.

Therefore, for the model \( y = x + v \) (\( y \) represents the image disturbed by noise, \( x \) represents the original clean image, and \( V \) represents the noise), it is a common practice to degenerate \( v \) into a Gaussian additive white noise subject to standard deviation [2]. The bigger the noise level \( \sigma \), the bigger the noise. The task of image denoising is to restore clean image \( x \) from image \( y \) with noise.

2.2. Convolutional neural network

Convolutional Neural Network (CNN) can automatically extract image features and has the characteristics of local perception and weight sharing. Such strong learning ability makes its performance indexes in the field of image noise reduction exceed many algorithms and obtains a large number of applications.

The convolutional neural network is a feed-forward supervised learning network, which is generally composed of a convolutional layer, activation layer, pooling layer, full connection layer, etc.

![Figure 1. Illustration of convolutional neural networks.](image)

The convolutional layer is the core structure of the convolutional neural network, which is used to extract the features of the input data. A two-dimensional convolution layer is commonly used to process image data, which can extract image edge, texture, local color, and other features. After local feature extraction, through multi-layer convolution superposition, a wider range of image features can be obtained. At the same time, due to weight sharing, all positions of the image can use the same statistical features, which greatly reduces the number of adjustable parameters, reduces the difficulty of training, and speeds up the running speed.
3

Figure 2. Convolution Computation Schematic

The activation layer introduces nonlinear transformation to the network, and improves the capability of the model's feature representation by making nonlinear mapping to the output features of the convolution layer [3], the commonly used activation functions include ReLU, Sigmoid, Tanh, etc. Compared with other activation functions, ReLU is simple to calculate and has the advantage of avoiding the problem of gradient disappearance and alleviating overfitting. ReLU is selected as the activation function of the network in the DnCNN model.

Figure 3. ReLU.

The convolutional neural network is an important classical structure in the deep learning model, which has been playing an active and effective role in computer vision processing tasks in recent years. As a typical low-level computer vision problem, the image denoising problem can also be processed by a convolutional neural network [4].

2.3. Residual learning and batch normalization

Batch normalization makes use of the mean and standard deviation of small batches to continuously adjust the intermediate output of the neural network, so as to make the value of the intermediate output of the whole neural network in each layer more stable. The addition of a batch normalization layer can ensure the balanced data distribution of feature vectors at each layer in the network structure, thus avoiding the disappearance of gradient, accelerating the convergence speed of model training, enhancing the generalization ability of the model, and making the training of deep network model easier and more stable.

In the process of image processing, with the increase of network depth, the training difficulty is enhanced, and the training accuracy begins to decline gradually. The residual network simplifies the learning objective by explicitly learning some residual mappings of stacked layers, thus making the training of the deep network easier [5]. At the same time, residual learning can improve the accuracy of image classification and target detection [6], improve the network performance, better learning target characteristics.

Zhang's study clearly shows that batch normalization and residual learning complement each other and that the combination of the two not only enables rapid and stable training but also generally leads to better denoising performance [7].

2.4. Channel attention mechanism

In recent years, the attention mechanism has been widely applied in various deep learning fields such as feature extraction, detection, classification, etc. The mechanism of attention is essentially similar to the mechanism of selective visual attention in humans [8], which can adjust the intermediate feature graph,
strengthen the important features and suppress the useless information on the basis of keeping the computational amount and complexity unchanged [9].

The channel attention mechanism generates a mask for the channel through the operation of the neural network and scores the value on the mask to evaluate the score of the current points that need attention, so as to strengthen the network's attention to the key information and improve the quality of the features generated by the network. Its core architecture is Shared MLP, and its basic operations are convolution and deconvolution.

After each convolution, the image will lose part of the information of the edge pixels and be blurred. Then deconvolution algorithm is used to reconstruct the noisy image, remove part of the noise information that cannot be reconstructed, extract more details and texture information, improve the segmentation accuracy of the network and make the photos clear. Because deconvolution network has such good feature learning ability, channel attention mechanism can effectively improve the representational performance of the convolutional neural network, and well realize the underlying computer vision image denoising function.

![Shared MLP](image)

Figure 4. Shared MLP.

3. DnCNN2 Model

3.1. DnCNN network architecture

The denoising convolutional neural network DnCNN algorithm framework is mainly composed of three different types of layers, which are displayed in three different colors. As shown in the figure, assuming the network depth is \(d\), (1) \(\text{Conv+ReLU}\): for the first layer, 64 filters with the size of \(3 \times 3 \times c\) are used to generate 64 feature maps. \(c\) represents the number of image channels, that is, when \(c=1\), it is a gray image, and when \(c=3\), it is a color image. Relu is used as an activation function to provide non-linearity for the network. (2) \(\text{Conv+BN+ReLU}\): for layers 2∼(d−1), 64 filters of size \(3 \times 3 \times 64\) are used, and batch normalization is added between the convolution layer and ReLU. (3) \(\text{Conv}\): in the last layer, only one convolution layer is used, and \(c\) filters of the size of \(3 \times 3 \times 64\) are used to output residual images [10].

![DnCNN network architecture](image)

Figure 5. The architecture of the proposed DnCNN network.

3.2. DnCNN2 model with the attention mechanism

The network structure designed in this paper is shown in the figure. In contrast to the DnCNN model, the attention mechanism was inserted in the DnCNN2 model before the penultimate layer. Shared MLP essentially uses the convolution kernel \([1,1]\) to deconvolve and set the stride as 1 and the bias as 0 to extract the feature information of the image. Similarly, the deconvolution layer with the size of convolution kernel 1 and step size 1 is used to restore the original size of the feature graph, so as to
enhance the sensitivity of the network to information features. Between the convolutional layer and the deconvolutional layer is the Ruluactivation function, which introduces nonlinear factors to the neuron and enables the neural network to approach any nonlinear function arbitrarily so that the neural network can use more nonlinear models. For the image \( y = x + v \) with noise, the residual learning is adopted to obtain the mapping \( R(y) = \tilde{v} \), and obtain the predicted residual image \( \tilde{x} = y - \tilde{v} \).

First of all, the improved DnCNN2 model will capture the noise of the noise-enhanced images. After residual learning and batch normalization processing, it will be sent into the channel attention mechanism for further feature extraction to enrich global semantic information and local detail information, and then the model will obtain noise feature images. Finally, the noise residual image is subtracted with the noise image, and the required denoising image can be output.

Figure 6. DnCNN2 model.

4. The Experimental Results

4.1. The experimental setup

In order to verify the effect of noise reduction on Gaussian noise with the known noise level, the training set of this model is BSD68, that is, 68 grayscale pictures with 481×321 or 321×481 pixels. Gaussian noise images with noise level \( \sigma = 15, 25 \) and 50 were selected to expand the data. The designed network depth is 18. L1 loss function and Adam optimizer are used in the training. The learning rate of the iterative process was initially set at 10^-4, the minimum batch size is 32. The model training lasted 50 times.

The data set used in the test was BSD68 and SET12.

4.2. Qualitative and quantitative assessment

To test the trained model, the evaluation of the denoising effect is mainly based on the following three indicators: 1. Human sensory judgment. 2. Peak signal to noise ratio (PSNR). 3. structural similarity SSIM.

PSNR represents the ratio of the maximum possible power of a signal to the power of the noise signal that affects its accuracy. The unit is dB. The larger the value, the smaller the distortion.

The calculation of SSIM is based on the comparison of brightness, contrast, and structural differences [11]. The value range is \([0, 1]\). The larger the numerical is, the smaller the image distortion is.

Table 1. The PSNR (dB) Results of Different Methods

| Methods | BM3D | EPLL | WNNM | DnCNN | DnCNN2 |
|---------|------|------|------|-------|--------|
| \( \sigma = 15 \) | 31.07 | 31.21 | 31.21 | 31.73 | 31.85 |
| \( \sigma = 25 \) | 28.57 | 28.68 | 28.83 | 29.23 | 30.48 |
| \( \sigma = 50 \) | 25.62 | 25.67 | 25.87 | 26.23 | 27.45 |

Table 2. The PSNR (dB) and SSIM Results of \( \sigma = 25 \)

| Methods | PSDR | SSDR | SSIM | DnCNN | DnCNN2 |
|---------|------|------|------|-------|--------|
| BSD68   | 29.23 | 30.48 |
| SSIM    | 0.8279 | 0.8290 |
| SET12   | 30.43 | 32.05 |
| SSIM    | 0.8617 | 0.8720 |
Figure 7. Denoising results of the image “parrot” with noise level 50. (a) Noisy / 25.00dB. (b) BM3D / 25.90dB. (c) WNNM / 26.14dB. (d) EPLL / 25.95dB. (e) DnCNN / 26.48dB. (f) DnCNN2 / 27.52dB.

Experimental results show that the DnCNN2 noise reduction model has a better performance compared with traditional algorithms such as BM3D and EPLL. Based on the comparison of DnCNN and DnCNN2 denoising models, there is almost no visual difference observed in the figure. After denoising, the noise of both models is significantly reduced, and basically clear edges, obvious textures, and more precise details can be restored, and the performance of solid color areas is very smooth. However, in terms of objective data of PSNR and structural similarity, compared with the original DnCNN model, the PSNR of BSD68 improved by 1.25dB in the DnCNN2 model with the attention mechanism, while the PSNR of SSIM improved by 0.0011. The PSNR of SET12 improved by 1.62dB and the SSIM improved by 0.0103, indicating a significant improvement.

5. Conclusion

Nowadays, image denoising has become a research direction with important applications in deep learning. DnCNN2, as an improved deep network residual learning image denoising model has the main advantages of DnCNN and ATTENTION mechanism. First, DnCNN2 has a deep hierarchical structure, which can effectively improve the flexibility and ability of image feature mining. Secondly, in the training process, ReLU activation function, batch normalization processing, and other methods were applied to solve many problems including gradient explosion and overfitting, which improved the noise reduction performance and accelerated the training rate. Third, image reconstruction under the attention mechanism makes the pixel range of noisy images more compact compared with the original image strengthens the key information in the image, and contributes to network convergence.

However, there are still limitations to this approach. Compared with DnCNN, although the computational efficiency of DnCNN2 has been improved, the whole algorithm still needs many iterations to obtain a good training model, and the rapidity and convergence of the whole algorithm are still not outstanding enough. And like most noise reduction models, the application in the actual scene is also very limited.

It is worth expecting that, with the deepening of the research on deep learning image denoising, more and more practical algorithm models will emerge.
References

[1] ZHANG K, ZUO W, CHEN Y, et al. Beyond a gaussian denoiser: Residual learning of deep CNN for image denoising[J]. IEEE Transactions on Image Processing, 2017, 26(7): 3142-3155.

[2] JIN Huazhong, LIU Yang, YE Zhiwei. An Improved Deep Network Residuals Learning Image Denoising Method [J]. Journal of Central China Normal University:Natural Sciences, 2020, 54(6): 949-955, 968.

[3] YAN Chuman, WANG Cheng. Development and Application of Convolutional Neural Network Model [J]. Journal of Frontiers of Computer Science and Technology, 2021, 15(1): 27-46 (in Chinese).

[4] LI Chuanpeng, QIN Pinle, ZHANG Jinjing. Research on image denoising based on the deep convolutional neural network [J]. Computer Engineering, 2017, 43(3): 253-260 (in Chinese).

[5] Michael N. Neural Networks and Deep Learning [EB/OL]. (2015) [2020-6-5]. http:// neuralnetworksanddeeplearning.com/.

[6] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition [C] // 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE Computer Society, 2016: 770-778.

[7] K. Zhang, W. Zuo, Y. Chen, D. Meng and L. Zhang, "Beyond a Gaussian Denoiser: Residual Learning of Deep CNN for Image Denoising," in IEEE Transactions on Image Processing, vol. 26, no. 7, pp. 3142-3155, July 2017, doi: 10.1109/TIP.2017.2662206.

[8] ZHANG Haitao, ZHANG Meng. SSD target detection algorithm with channel attention mechanism [J]. Computer Engineering, 2020, 46(8): 264-270.

[9] YUAN Jiajie, ZHANG Ling, CHEN Yunhua. Deep neural network image recognition based on attention convolution module [J]. Computer Engineering and Applications, 2019, 55(008): 9-16.

[10] LIU Liping, QIAO Lele, JIANG Liucheng. Overview of image denoising methods [J/OL]. Computer Science and Exploration: 1-16 [2021-03-25 14:52] (in Chinese).

[11] ZHANG Zhonghan. (in Chinese) Research on Image Denoising Algorithm Based on U-NET and Its Performance Optimization [D]. Jilin University.