

# Mixed Noise Removal by Residual Learning of Deep CNN

Kang Yang<sup>1</sup>, Jielin Jiang<sup>1,2,\*</sup> and Zhaoqing Pan<sup>1,2</sup>

<sup>1</sup>School of Computer and Software, Jiangsu Engineering Center of Network Monitoring, Nanjing University of Information Science and Technology, Nanjing, 210044, China

<sup>2</sup>Country Jiangsu Collaborative Innovation Center of Atmospheric Environment and Equipment Technology, Nanjing University of Information Science and Technology, Nanjing, 210044, China

\*Corresponding Author: Jielin Jiang. Email: jiangjielin2008@163.com Received: 06 December 2019; Accepted: 06 December 2019

Abstract: Due to the huge difference of noise distribution, the result of a mixture of multiple noises becomes very complicated. Under normal circumstances, the most common type of mixed noise is to add impulse noise (IN) and then white Gaussian noise (AWGN). From the reduction of cascaded IN and AWGN to the latest sparse representation, a great deal of methods has been proposed to reduce this form of mixed noise. However, when the mixed noise is very strong, most methods often produce a lot of artifacts. In order to solve the above problems, we propose a method based on residual learning for the removal of AWGN-IN noise in this paper. By training, our model can obtain stable nonlinear mapping from the images with mixed noise to the clean images. After a series of experiments under different noise settings, the results show that our method is obviously better than the traditional sparse representation and patch based method. Meanwhile, the time of model training and image denoising is greatly reduced.

Keywords: Mixed noise denoising; residual learning; deep convolutional neural

## **1** Introduction

In the field of computer vision, image denoising is often the first step and the basis of all subsequent research operations. Noise may appear in the process of image shooting and transmission, which is often inevitable. Noise can be defined as "unpredictable" in theory, which can only be recognized by probability statistics method [1]. When faced with the problem of image denoising, we often hope to rely on the prior distribution of noise. When the image quality is degraded due to the interference and influence of various noises, it will have adverse effects on the subsequent image processing and image visual effect. There are many kinds of image noise in daily life, including electrical noise, mechanical noise, channel noise and other noises. Therefore, in order to suppress noise, improve image quality and facilitate higher-level processing, it is necessary to denoise the image to preserve the image details, texture, edging structure.

The most common type of mixed noise is to add impulse noise (IN) and then white Gaussian noise (AWGN) [2]. AWGN is the most basic noise and interference model and its amplitude distribution obeys Gaussian distribution. If the probability distribution of white noise values obeys Gaussian distribution, such noise is called Gaussian white noise [3]. IN is a general term for discrete noise in communication. It is a kind of random white or black dot, which may be black pixels in the bright area or white pixels in the dark area (or both). Impulse noise may be caused by sudden strong interference in the image signal, analog digital converter, or error in bit transmission [4]. For example, a failed sensor results in a minimum pixel value, and a saturated sensor results in a maximum pixel value.



First, the methods of eliminating the disturbance of Impulse noise should be concerned. Nonlinear filters (such as median filters [5]) have been widely used to remove IN. However, by using this simple filtering technique to destroy the serious noise, the local structure of the image will be destroyed. Among them, weighted median filter [6], central weighted median filter [7] and multi-state median filter [8] may operate pixels without noise as pixels with noise. In this way, the details of the image will be erased. Some filters operate only on damaged pixels in the image, so a method of detecting noisy and noiseless pixels is needed. Denoising methods using this idea include switching MF [9], adaptive MF (AMF) [10] and adaptive CWMF (ACWMF) [11].

Then, the removal of AWGN should be paid attention to. In the area of image denoising, AWGN is a more general noise, and more research is applied to AWGN. Commonly used methods include wavelet transform [12], fourier transform, sparse representation based on over-complete dictionary, transformation based on multi-scale geometric analysis. All of these methods use the local information of the image, that is, denoising is achieved by weighted average of all pixels or parts of pixels in the neighborhood of pixel points. In this case, however, the edge is not retained, resulting in edge displacement, edge disappearance or even phantom edge. In 2006, basing on image contains large amount of self-similar structure. Wang et al. [13] proposed the characteristics of non-local average denoising algorithm. This algorithm uses the weighted average idea to estimate the value of the pixels to be processed, and the weight is calculated by the similarity between the two image blocks, which calculated by the Euclidean distance weighted by Gauss. NLM makes full use of a great deal of redundant information in natural images to better maintain the details and textures of images. In 2007, Dabov et al. [14] came up with BM3D, the method based on image Block in between, with similar structure of the image Block for 3D transform domain Filtering wave. In 2009, Tadize was put forward based on principal component analysis method, the method using PCA image neighborhood vector projection to the lowdimensional subspace, and then use the distance meter to calculate the similarity value of subspace [15]. In 2016, Cai et al. [16] proposed a method based on candidate set selection, which firstly searched for image blocks with similar gray distribution to form candidate sets. In 2017, Nguyen et al. [17] improved the James-Stein center pixel weight estimation method.

At last, when it comes to mixing noise, it is more troublesome because the characteristics of these two kinds of noise are completely different. At present, some methods are used in the denoising of mixed images. Some methods originally used to remove IN noise are directly applied to mixed noise, but the effect is often not very ideal, there will be artifacts in the input. In order to detect in more effectively, there are many new methods based on BF. Trigonal filter [18] is to add the absolute difference of order (Road) statistics to BF. The method of switching bilateral filter (SBF) is to continuously switch between two different kinds of noise removal, so as to achieve the overall denoising. But the premise is that the classification of noise needs to be accurate, so the reference median is used to determine whether the current pixel is a noisy pixel or not. If the absolute value between the reference median and the current pixel is large, the pixel with noise is determined. Cai et al. [19] proposed an improved two-stage method. Firstly, the candidate impulse noise is removed, and then the result is used to denoise and blur, so as to avoid errors in the damaged pixels and achieve an effective removal effect. Xiao et al. [20] proposed 11-10 minimization method for mixed noise removal, which achieved the most advanced noise reduction results. However, due to the complicated calculation, it still needs to be improved. Liu et al. [21] proposed a new weighted regularization term which is different from the previous one and has better efficiency in optimization. At the same time, sparse coding and dictionary learning are used, and the improved K-SVD is used for processing. Mao et al. [22] proposed to apply the deep convolutional codec network to image denoising. In the coding layer of the convolutional network, the encoding layer is immediately after the decoding layer by jumping every two layers to obtain the mapping from the damaged image to the noisefree one. Jiang et al. [23] proposed a weighted coding sparse nonlocal regularization (WESNR) method. He improved the previous two-stage method, without separate detection steps. Instead, remove all the noise at once. Firstly, the contaminated image is replaced by AMF, which makes the noise distribution from the previous confusion to approximate Gaussian distribution, which can be more effective for the

subsequent steps. Zhang et al. [24] used deep learning correlation method to denoise the image. For the sake of all kinds of problems due to the increasing of network depth, the l2 norm of output and noise was taken as a loss function to train the network, which could effectively remove uniform Gaussian noise. However, for non-uniform noise, the effect is decreased, and the network can be regarded as a residual learning process, which can be better trained. FFDNet uses noise level estimation as the input of the network, which can deal with more complex noises, such as noise with different noise levels and noise with spatial variation. The network USES orthogonal matrix to initialize the network parameters, which makes the network training more efficient. The input image is sampled under multiple sub-images as the network input, and the output sub-images are sampled up to obtain the final output.

All the above methods have certain effects, but when the noise value is unknown or very large, the effect will be very poor. Therefore, this paper provides a method of mixed noise removal by the use of residual learning of deep CNN, which achieves the effect of blind Noise Removal by using Residual Learning. It has good effect when the noise value is low and very good effect when the noise value is high. At the same time, compared with our methods, the cost of model training and testing of the previous methods is much higher than ours. When the noise value is high, compared with other methods before, the effect is significantly improved. Meanwhile, when the noise is low, the effect of this method is good as well.

## 2 Residual Learning to Remove Mixed Noise

#### 2.1 Residual Learning

At the beginning, residual learning mainly serves the training of network when the depth of net is getting deeper and deeper. If the later layers of the deep network are identical mappings, then the model will degenerate into a shallow network. That is, the deeper a network is, the less effective it will be in training the network on the training set, which is why we are sometimes reluctant to deepen the layers. In the past, we thought that increasing depth can improve performance. Because the two networks with different depth, the low depth network can be directly put into the high depth network, and the subsequent layer will do identity mapping. However, when the depth increases to a certain extent, the effect does not change or even rapidly decreases.

However, if we define the network as S(x) = O(x) + x, It can be rewritten to learning a residual value as O(x) = S(x) - x (see Fig 1). When O(x) = 0, you're forming an identity map where S(x) = x. Moreover, it is definitely easier to fit residuals. Moreover, improved accuracy for image classification and target detection has been achieved [25].



Figure 1: Residual learning

In this paper, residual learning formula is used in our proposed model. However, our model does not use too many residual units, only one is used to remove mixed noise.

## 2.2 Network Architecture

Generally speaking, the model generated by CNN has complicated processing before and after training, which is generally considered to be non-operational. So in many cases, the input is preprocessed to improve the capability of the neural network. In the practices of various literatures, almost all of them use preprocessing method to firstly reduce the noise of the image destroyed by the mixture of different noise, and then proceed to the next step. Therefore, we preprocess in the first step of the model.

Preprocessing generally uses a combination of various filters applicable to IN to select images according to the type of destruction. At present, there are many filters that can achieve good results, such as MF, AMF, CWMF, ACWMF and so on. Their corresponding noise situation is also different. AMF is very adequate for salt and pepper noise, while MF is also apt for the noise above plus random noise. According to different failure types, different filter combinations are selected during pretreatment. This is the preprocessing part, the first step in the model (Fig. 2).

After the first step of preprocessing, the input of the model is still an image damaged by noise y = x + v. General discriminant denoising models ultimately want to get a noise-free image, and the mapping function it learns is F(y) = x. According to the principle of residual learning given above, our model can obtain the residual mapping  $S(y) \approx v$ , then the residual of noisy image and clean one can be received in this way. By subtracting the original noise image from the residual image, that is, x = y - S(y), a clean image can be obtained.

$$\ell(\Theta) = \frac{1}{2N} \sum_{i=1}^{N} \|S(y_i; \Theta) - (y_i - x_i)\|_F^2$$
(1)

The above loss function is used to train the parameter  $\Theta$  needed by the model. In this function,  $\{(y_i, x_i)\}_{i=1}^N$  is N pairs of noisy and cleaning patches.



Figure 2: Network architecture of proposed model

Network Architecture: We set the depth of proposed model to 17, and different of layers are adopted in our model, as shown in Fig. 2. After pre-processing, we used one Conv plus ReLu layer to get a multi-channel result. That's because we use 64 convolution kernels, and they are all the same size of  $3 \times 3$ . So we can get 64 feature graphs in this layer. In order to make our results more robust, the activation function ReLU is used to make the results more nonlinear. The next 15 layers are all the same: Conv + BN + ReLU. 64 convolution kernels are used in each layer. Their size are  $3 \times 3 \times 64$ . BN is added to the layers to let the active input distribution of each hidden layer node be fixed. At last, we only one convolution kernels of size  $3 \times 3 \times 64$  to obtain the final feature map.

In conclusion, our proposed model has the following main characteristics: We used preprocessing technology to make the noise distribution more traceable. The heavy tail produced by the original mixed noise distribution is eliminated. Then, we used residual expression to calculate S(y). BN is adopted to let the active input distribution of each hidden layer node be fixed, so that training efficiency has been improved. By using linear rectifying function, we can make the training result have better nonlinear relation, which makes our method more universal. Finally, our algorithm can separate the reasonable approximation of the noise from the noise image, and the prediction of the original image can be obtained after the subtraction operation of the approximation of the noise.

## **3** Experiments

## 3.1 Training Data

For noise images with different noise levels, we use the same data set for training. These are 400 color images from BSD, and we converted them to grayscale. They are going to be  $180 \times 180$ . The reason for not using a larger data set is that a larger data set does not have a significant effect boost. We slice these pictures, and perform flipping, folding and other operations to increase the amount of data. The size of the slices is  $40 \times 40$ . At last, we can get  $128 \times 1$ , 600 patches to put into the model.

Referring to the previous practice, we used 12 typical images widely used in image denoising field to test our performance. Moreover, these images do not exist in the data set of 400 used to train the model described above. As a result, the test results are more convincing.

## 3.2 Parameter Setting

Due to the use of residuals, we can set a relatively deep network structure, but too many layers will affect the training speed, so the depth is set to 17. Since the stochastic gradient descent method is used as the optimizer, its parameter setting is crucial for training. The number of iterations is set to 180, the initial learning rate is 1E-3, and it becomes one fifth of the current learning rate in the 30th, 60th, and 90th iteration. The weight attenuation is 0.0001, the momentum is 0.9, and the minimum batch size is 128.

#### 3.3 Results

We have done a lot of model training and testing, with their results to prove the effectiveness of our algorithm. Pytorch is used to train the models, which is a widely used deep learning framework. Our training environment is Pycharm on a DELL workstation with Intel i7-9700k CPU 3.60 GHz and Nvidia GTX 2080Ti GPU.

$\sigma=30$ and $\rho=0.2$	WESNR	11–10	Proposed Method	$\sigma=30$ and $\rho=0.3$	WESNR	11–10	Proposed Method
C.man	23.08	26.66	27.51	C.man	21.74	25.84	25.99
House	29.88	30.22	31.28	House	29.35	29.41	30.14
Peppers	26.31	27.48	28.06	Peppers	25.27	26.63	26.77
Starfish	25.39	26.27	27.28	Starfish	24.95	25.49	26.16
Monar.	25.19	26.52	28.02	Monar.	23.72	25.73	26.62
Airpl.	24.68	25.98	26.65	Airpl.	23.27	25.31	25.67
Parrot	21.04	26.13	26.94	Parrot	20.27	25.32	25.66
Lena	29.54	29.72	30.74	Lena	29.19	29.03	29.81
Barbara	26.04	27.32	27.35	Barbara	25.34	26.61	26.01
Boat	27.16	27.54	28.36	Boat	26.65	27.06	27.45
Man	27.35	27.55	28.28	Man	27.06	27.14	27.51
Couple	26.71	27.18	28.26	Couple	26.37	26.65	27.35
Average	23.93	27.38	28.24	Average	25.26	26.68	27.09

 Table 1: The PSNR results obtained by WESNR, 11
 10 and ours under low noise level

Tab. 1 lists the PSNR results of 12 typical test images by different methods under low noise level. It is obvious and intuitive that the results are presented to us, that our method has better results and better performance than the competitive methods. Compared to WESNR, the proposed method that we came up with has a notable PSNR gain of about 4.31 dB and 1.83 dB. Compared to the method 11–10, our approach is much better than it. PSNR increased by about 0.86 dB and 0.41 dB. When the noise level is higher, the advantage of our model is better reflected.

$\sigma=40$ and $\rho=0.2$	WESNR	11–10	Proposed Method	$\sigma=40$ and $\rho=0.3$	WESNR	11-10	Proposed Method
C.man	20.47	25.54	26.62	C.man	19.38	24.58	25.33
House	26.83	28.27	29.96	House	25.13	27.57	28.62
Peppers	23.33	26.31	26.88	Peppers	22.19	25.32	25.83
Starfish	22.74	24.77	25.77	Starfish	21.47	24.07	24.87
Monar.	21.21	25.22	26.53	Monar.	19.56	24.37	25.46
Airpl.	20.81	24.67	25.69	Airpl.	19.53	23.88	24.63
Parrot	18.63	25.09	26.09	Parrot	17.49	24.26	24.84
Lena	27.16	28.21	29.45	Lena	26.09	27.57	28.32
Barbara	22.51	25.78	25.98	Barbara	20.99	25.02	24.71
Boat	24.68	26.28	27.26	Boat	23.43	25.72	26.42
Man	25.29	26.39	27.2	Man	24.41	25.88	26.41
Couple	24.68	25.81	27.01	Couple	23.58	25.33	26.02
Average	23.19	26.02	27.04	Average	21.93	25.29	25.96

Table 2: The PSNR results obtained by WESNR, 11\_10 and ours under high noise level

Tab. 2 lists the PSNR results of 12 typical test images by different methods under high noise level. Again, the results can be clearly presented. When the noise level is high, the two competing methods are not efficient. By contrast, our proposed methods still have a notable PSNR.



(a) Mixed Noise



(b) 11–10



(c) WESNR



(d) Our Method **Figure 3:** C.man  $\sigma$  = 40 and  $\rho$  = 0.2

6







**Figure 6:** Starfish  $\sigma = 40$  and  $\rho = 0.3$ 

Figs. 3-6 show the specific results of each algorithm. It is not difficult to find that although WESNR is good in numerical value, it is not good at the edge of the image. Some of the edges of the texture were erased. By comparison, our method is better for details retention, not too many edge details are erased to be very smooth. At the same time, it gives people a good visual feeling.

Not only information of the edge is preserved more completely. We can see that in the starfish image, our result also has more reservation for non-edge details. As you can see, only our method can keep the spots and shadows in the background information.

In the image man, the metal color of the camera bracket is not good in WESNR. The metal color is mixed with the black background, and the edge is even distorted. On the contrary, our method has a good restoration effect for the details, and has a better visual experience.

#### **4** Conclusion

In this paper, we apply the idea of residuals to the removal of image mixed noise. A deep learning method based on residual is proposed to remove the mixed noise of images. Firstly, different filters are used to preprocess the irregular noise image to AWGN noise. Then batch normalization and ReLU are applied to the training model to improve the training efficiency. Moreover, the idea of residual is applied to the training, and the loss function is also set according to the residual, which improves the de-manic performance. And finally, after a lot of experimentation, it turns out that our approach is obviously due to the traditional approach. At the same time, the time of training and desiccation has been greatly shortened.

Acknowledgement: We gratefully acknowledge the support from NVIDIA Corporation for providing us the 2080Ti GPU that was used in this research. I am deeply thanking to my tutors and my parents for their continuous support and encouragement.

**Funding Statement:** This work was supported in part by the National Natural Science Foundation of China under Grant 61601235, in part by the Natural Science Foundation of Jiangsu Province of China under Grant BK20160972.

**Conflicts of Interest:** We declare that we do not have any commercial or associative interest that represents a conflict of interest in connection with the work submitted.

#### References

- T. F. Rabie, "Adaptive hybrid mean and median filtering of high-ISO long-exposure sensor noise for digital photography," *Journal of Electronic Imaging*, vol. 13, no. 2, pp. 264–278, 2004.
- [2] J. Delon and A. Desolneux, "A patch-based approach for removing impulse or mixed gaussian-impulse noise," SIAM Journal on Imaging Sciences, vol. 6, no. 2, pp. 1140–1174, 2013.
- [3] Y. L. Liu, J. Wang and X. Chen, "A robust and fast non-local means algorithm for image de-noising," *Journal of Computer Science and Technology*, vol. 23, no. 2, pp. 270–279, 2008.
- [4] G. Pok, J. C. Liu and A. S. Nair, "Selective removal of impulse noise based on homogeneity level information," *IEEE Transactions on Image Processing*, vol. 12, no. 1, pp. 85–92, 2003.
- [5] G. Sicuranza and G. Ramponi, "Adaptive nonlinear digital filters using distributed arithmetic," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 34, no. 3, pp. 518–526, 1986.
- [6] L. Jin and D. Li, "A switching vector median filter based on the CIELAB color space for color image restoration," *Signal Processing*, vol. 87, no. 6, pp. 1345–1354, 2007.
- [7] S. J. Ko and Y. H. Lee, "Center weighted median filters and their applications to image enhancement," *IEEE Transactions on Circuits and Systems*, vol. 38, no. 9, pp. 984–993, 1994.
- [8] A. Nieminen, P. Heinonen and Y. Neuvo, "A New Class of Detail-Preserving Filters for Image Processing," IEEE Transactions on Pattern Analysis & Machine Intelligence, vol. 9, no. 1, pp. 74–90, 1987.
- [9] S. Zhang and M. A. Karim, "A new impulse detector for switching median filters," *IEEE Signal Processing letters*, vol. 9, no. 11, pp. 360–363, 2002.
- [10] R. Garnett, "A universal noise removal algorithm with an impulse detector," *IEEE Transactions on Image Processing*, vol. 14, no. 11, pp. 1747–1754, 2005.
- [11] C. H. Lin, J. S. Tsai and C. T. Chiu, "Switching bilateral filter with a texture/noise detector for universal noise removal," *IEEE Transactions on Image Processing*, vol. 19, no. 9, pp. 2307–2320, 2010.
- [12] Y. Zhang, "Wavelet-based Bayesian fusion of multispectral and hyper spectral images using Gaussian scale mixture model," *International Journal of Image and Data Fusion*, vol. 3, no. 1, pp. 23–37, 2012.
- [13] Y. L. Liu, J. Wang and X. Chen, "A robust and fast non-local means algorithm for image de-noising," *Journal of Computer Science and Technology*, vol. 23, no. 2, pp. 270–279, 2008.
- [14] K. Dabov, A. Foi and V. Katkovnik, "Image denoising by sparse 3-D transform-domain collaborative filtering," *IEEE Transactions on Image Processing*, vol. 16, no. 8, pp. 2080–2095, 2007.
- [15] T. Tasdizen, "Principal neighborhood dictionaries for nonlocal means image denoising," IEEE Transactions on Image Processing, vol. 18, no. 12, pp. 2649–2660, 2009.
- [16] W. F. Sun and Y. H. Peng, "An improved non-local means de-noising approach," Dianzi Xuebao (Acta Electronica Sinica), vol. 38, no. 4, pp. 923–928, 2010.
- [17] L. Bo, J. Lv and X. Luo, "A novel and fast nonlocal means denoising algorithm using a structure tensor," *Journal of Supercomputing*, vol. 75, no. 2, pp. 770–782, 2019.
- [18] R. Garnett, "A universal noise removal algorithm with an impulse detector," *IEEE Transactions on Image Processing*, vol. 14, no. 11, pp. 1747–1754, 2005.
- [19] J. F. Cai, R. H. Chan and M. Nikolova, "Two-phase approach for deblurring images corrupted by impulse plus Gaussian noise," *Inverse Problems and Imaging*, vol. 2, no. 2, pp. 187–204, 2008.
- [20] Y. Xiao, T. Zeng and J. Yu, "Restoration of images corrupted by mixed Gaussian-impulse noise via 11–10 minimization," *Pattern Recognition*, vol. 44, no. 8, pp. 1708–1720, 2011.
- [21] J. Liu, "A weighted dictionary learning model for denoising images corrupted by mixed noise," *IEEE Transactions on Image Processing*, vol. 22, no. 3, pp. 1108–1220, 2012.

- [22] V. Badrinarayanan, A. Kendall and R. Cipolla, "Segnet: a deep convolutional encoder-decoder architecture for image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 12, pp. 2481–3155, 2017.
- [23] J. L. Jiang, L. Zhang and J. Yang, "Mixed noise removal by weighted encoding with sparse nonlocal regularization," *IEEE Transactions on Image Processing*, vol. 23, no. 6, pp. 2651–2662, 2014.
- [24] K. Zhang, W. Zuo and L. Zhang, "FFDNet: toward a fast and flexible solution for CNN-based image denoising," *IEEE Transactions on Image Processing*, vol. 27, no. 9, pp. 4608–4622, 2018.
- [25] S. Wu, S. Zhong and Y. Liu, "Deep residual learning for image steganalysis," *Multimedia Tools and Applications*, vol. 77, no. 9, pp. 10437–10453, 2018.
- [26] K. Zhang, W. Zuo and Y. Chen, "Beyond a gaussian denoiser: Residual learning of deep CNN for image denoising," *IEEE Transactions on Image Processing*, vol. 26, no. 7, pp. 3142–3155, 2017.