Research Progress in Image Denoising Algorithms Based on Deep Learning

Cai Ruikai
College of Electronic and Information Engineering Electronic Information Engineering, Shenzhen university ,518061

Abstract: Images constitute the main source of information to people. Digital images are vital means of obtaining, processing, analyzing, and sharing information in the era of information. Now they have been deeply incorporated into every aspect of people’s production and life, generating considerable social and economic benefits. Thus improving the quality of images and reducing the negative impact of image noises to subsequent image processing have been two important research topics. Image processing technology has been combined with such research fields as cognitive psychology, machine learning, machine vision, and deep learning in recent years, which lead to an unprecedented development level and breakthroughs. Therefore, the study of image denoising technology has profound theoretical significance and promising prospects in practical application. The paper mainly discusses the development course of applying deep learning technology to the image denoising field. Meanwhile, it introduces various classical denoising algorithms and focuses on the thoughts, advantages, and disadvantages of each algorithm. Furthermore, it discusses the challenges faced by deep learning in the denoising field and puts forward possible solutions.

1. Introduction
The vision system has been the main source of information to humans, which means images are important information carriers. With the development of digitalization, various imaging devices provide a large number of digital images that can be processed and analyzed easily. Nowadays, most technology fields are related to digital images by varying degrees. Meanwhile, images not only constitute the main source of information, but are also important types of data for machines to obtain, process, analyze, and learn. Therefore, an in-depth study of the image science plays a large role in helping humans to understand and reconstruct the world better.

One of the major research domains of image science is image processing. It mainly processes, exchanges, and improves the visual effect of images to lay a foundation for subsequent application at a medium-to-high level, such as recognition, analysis, and decision-making. In the application after image processing, the quality of images is an important precondition for all studies. In fact, the quality of images in imaging, storage, and DAC tend to decline due to such factors as the defects of sensor hardware, the limitation of broadband, environmental noises, lossy compression and the accuracy of DAC. It brings significant disturbance and deviations to subsequent image processing, such as image segmentation and target recognition. As a result, image understanding, prediction, and analysis will face difficulties. In particular, the quality of images directly affects the accuracy and reliability of recognition results in the most popular AI field, such as WIT 120 and driverless technology. It thus determines the success or failure of such projects. Therefore, studying the methods for denoising and recovering high-quality images from damaged images is the most fundamental and important task of image processing. The study has tremendous theoretical significance and high application values.
Although this field has been studied for more than half a century, it is still worthy of further explorations. Various techniques, algorithms, and processing thoughts have emerged endlessly in past explorations. Traditional processing methods include the image denoising methods based on spatial arithmetic mean filtering, Gaussian filtering, median filtering, maximum filtering, and minimum filtering; image denoising methods based on wavelet transformation domain; image denoising methods based on BM3D that combines the spatial and transformation domain thinking. Traditional denoising methods generally lead to the loss of image details. Meanwhile, their denoising performances and the complexity of algorithms are unsatisfactory. Thanks to the significant improvement in the performance of graphics cards, there reoccurs the wave of reviving artificial neural networks. Many scholars have applied this algorithm to image denoising, such as MLP, stacked sparse autoencoder, and CNN, which all achieved satisfactory results. Thanks to its high learning capacity, DNN has been widely applied to processing images, speech, and other signals, which achieved significant achievements in the image field. The application of deep learning to image denoising also proved its huge potential. On the other hand, the application of deep learning to image denoising is a cutting-edge direction, whose theories and technology are not mature. Further studies should be carried out to find how to better apply the advantages of deep learning to the denoising field.

2. Traditional Methods

In traditional methods, scholars mainly study image denoising from two perspectives, including the spatial domain and the transformation domain.

2.1 Spatial Domain

2.1.1 Linear Filtering

The arithmetic mean filtering \[^1\] replaces the pixel value at the center of the pixel block on the image with the average gray level of other pixel values in the neighborhood for noise reduction and image smoothing. This method has intuitive concepts and low calculation complexity. Conversely, it has some disadvantages: Firstly, the edge details of the image may be lost during noise reduction; secondly, the inadequate number of gray levels may lead to the pseudo-contour effect.

The Gaussian filtering \[^2\][^3\] is similar to the arithmetic mean filtering. The only difference is that the former method determines the value of each pixel point through the weighted average value of the pixel point and other pixel values in the neighborhood. Both methods process digital signals through simultaneous smoothing. There is no striking difference. Therefore, the Gaussian filtering has similar defects to those of the arithmetic mean filtering.

2.1.2 Nonlinear Filtering

Turkey was the first to put forward the concept of median filtering in 1974 to solve the problems of discrete data smoothing \[^4\]. Shortly after its proposal, this new thinking was introduced into the image denoising field and create the standard median denoising algorithm \[^5\][^6\]. Median filtering is non-linear spatial filtering, whose thinking is to rank the gray levels of all other pixels in the neighborhood centered on the pending element and assign the median value to the pending element. In general, the template and window refer to selected neighborhoods. Different templates and sizes have different impacts on the denoising effect. It is thus necessary to adjust the size and shape to obtain the best denoising effect. This algorithm protects the edge characteristics of images and restricts the deviate Gaussian distribution and even incompletely independent signals. Thus the algorithm has better denoising effect on salt & pepper noises. Conversely, its capacity for processing random noises is slighter weaker than that of average mean filtering. Meanwhile, it may lead to the loss of acute angles and line segments and destroy the image’s geometrical structure when filtering noises.
2.2 Transformation Domain
Wavelet transform filtering [7]-[10] shifts the image from the spatial domain into the wavelet domain for processing. The wavelet transform domain retains the positional information of the corresponding spatial domain. When multi-scale wavelet transformation is conducted, the component of each resolution rate corresponds to the position of the spatial domain. It is more favorable for abstracting characteristics and denoising. When the wavelet transform is conducted, different wavelet bases can be selected flexibly to process different types of images. Meanwhile, different threshold values should be set according to the strength of image noises to filter the wavelet domain and eliminate the interference of noises. After the wavelet coefficient filtered by the threshold value is obtained, the sparse coefficient is input into the spatial domain through wavelet reverse transformation to obtain denoised images. Although the noise-reduction through transformation domain improves the visual effect of image denoising, the selection of threshold values for wavelets is generally the biggest difficulty of denoising. If the threshold value is high, the image will have ambiguous details. If the threshold value is small, the noises cannot be eliminated entirely.

2.3 Thinking Combining Spatial Domain and Transformation Domain
BM3D matches similar blocks of an image and overlaps the image blocks with similar conditions into a 3-D matrix. A 3-D matrix contains a large amount of redundant information. Thus the 3-D matrix undergoes the 3-D frequency domain decomposition to achieve frequency domain coefficients with better sparsity. BM3D combines image sparsity and non-local characteristics, which achieves good results. It has a significantly better effect of restoring the veins of an image than other traditional methods. Regarding the effect of restoring simple veins such as lines, it can achieve a consistent visual effect with the original image when the noise strength is low. Meanwhile, BM3D conducts threshold filtering processing for coefficients in the transformation domain. If larger threshold values are selected to process images with high noises, it may cause truncation errors and artifacts in the de-noised image, which affects the accuracy of denoising results.

3. Deep Learning

3.1 Multi-layer Perception
MLP (Multi-layer Perception) is a non-linear function that maps the vector input to the vector output through multiple hidden layers. It has a strong non-linear capacity and can approximate any function. Burger et al. applied MLP to image denoising in 2012. The core thinking of MLP is: to directly learn the mapping relationship between the original image and the noisy image. Training process: Firstly, noises are added to a noiseless image to produce a noisy image. Next, the noisy image is input into MLP. Then the differences between the output image and original noiseless image (such as the mean square error) are taken as the loss and cost. Subsequently, the parameter of MLP is updated through the reverse gradient propagation algorithm to minimize the cost & loss and obtain a high-quality denoising image. It is too complex to directly learn the calculation for the noising mapping function of a whole image. Thus an image is generally divided into several patches for study and denoising in the learning process. The size of each patch is crucial to the final quality of the denoising function. If the size is too small, the function is easy to learn yet has a poor denoising effect. If the size is too large, it can get better denoising performances yet a harder-to-train model.

This method applies to the training of GPU. Although it takes a longer time than BM3D, it is much faster than other methods that need a learning dictionary (such as KSVD or NLSC). Its denoising performances are affected by numerous factors, including the size of the training set, the structure of MLP, and the patch size. The adjustment of such allocations is severely reliant on artificial experience. Also, its denoising performances for processing the images with many similar structures may be substantially lowered.
3.2 Convolutional Neural Network
CNN (Convolutional Neural Network) is a variant of MLP (Multi-layer Perception) and plays an important role in current studies of deep learning. CNN is more similar to the weight-sharing network structure of biological neural networks. It can lower the complexity of the network model and decrease weights. Furthermore, images can be used as network input directly, which avoids the complexity of data reconstruction in characteristic abstraction and classification. CNN thus has unique advantages in image processing. When CNN is applied to image denoising, convolutional layers generally need to be stacked to construct a deep network structure to improve the learning capacity and denoising performances of the network. For one thing, it may lead to a substantial increase in parameters, enhance the difficulty of training, and lead to over-fitting. As the network depth increases, the gradient dispersion becomes non-negligible, and model optimization becomes more difficult. Also, the learning rate of each convolutional layer should be constantly adjusted in the training process to achieve better fitting for the CNN model. It thus increases the difficulty of overall network adjustment and optimization.

According to the problems above, the algorithm has unique advantages and development potential in the image denoising field. It is because CNN has the following characteristics of neural network and the skills of deep learning: (1) local receptive field—The receptive field represents the calculation region where the pixel points of the output image correspond to the original input image. The higher the receptive field is, the bigger the search area for recovering image pixels is. In other words, the size of the convolution kernel represents the capacity of learning the noise characteristics of a wide range. Meanwhile, the stacking of convolution kernels expands the perceptive field. (2) Batch normalization: The image characteristics data abstracted from each convolution layer in training undergo normalized processing. It enables each network to learn the same distribution of characteristic data and improves the training efficiency of the network. (3) Multi-scale characteristic abstraction: Different image characteristics can be abstracted from different convolution kernels. Thus the convolution kernels of different sizes should be used to scan input noisy images at the same time and abstract different noise characteristics. This technology considerably accelerates the speed of actual training and makes it easier for the model to converge. (4) Residue learning [18]: It aims to directly learn from the tiny differences between noises and the original image. In other words, learning from noises requires a smaller dimension of noise characteristics compared to learning from image contents. Meanwhile, the deep network can be trained more easily to effectively prevent the gradient disappearance of the deep network in back-propagation. (5) Dropout: The technology offers a probability in the training of the neural network and temporarily sets some network nodes to zero, so that their roles in the network become ineffective. Also, it makes the network more adaptable and effectively avoids over-fitting.

Due to the characteristics above, the possibility of improving denoising performances by training the deep network is substantially raised. Meanwhile, it somehow alleviates the issues of degradation.

3.3 Generative Adversarial Network

![Fig.1. Schematic diagram of GAN denoising network](image)

Clean image

Discriminator

Generator

Noisy image

De-noised image
GAN (Generative Adversarial Network) was first put forward by Goodfellow. It is a non-supervision learning method. The original GAN was used to learn from the sample images in the image library and generate images similar to such samples. After some scholars solved the difficulty of convergence, GAN is widely applied to various fields of computer vision and has achieved achievements in the image denoising field.

When GAN is applied to image denoising, the generator is used for image denoising. Meanwhile, another network, namely the discriminator, is used to adjust the structure of the generator. The discriminator is used to differentiate the noisy image and de-noised image. The output is a probability value that suggests the similarity between the de-noised image output by the generator and the real image. Next, the similarity is taken as a part of the generator’s loss function to guide the generator in subsequent training. Eventually, the improved de-noising image obtained by the generator is used to train the discriminator. Such is the so-called adversarial training. The discriminator differentiates the insignificantly improved de-noised image and the noiseless image through adversarial training to make the vein details and contrast ratios of both images more similar to each other. The de-noised image retains more high-frequency characteristics, which thus improves its visual image.

GAN does not patch up sample image blocks into the output image by memorizing these blocks through the network. Instead, it learns from the image characteristics of the sample image blocks stored in the database and reconstructs the image through such characteristics. Generally speaking, the pre-training conditions for the discriminator are strict: If the discriminator is poorly trained, it fails to tell the differences between the produced image and the real image. As a result, the loss function that contains the discriminator value may be inaccurate in training for the generator and thus fails to guide image generation. If the discriminator is over-trained, the generator cannot be optimized.

3.4 Stacked Sparse Autoencoder
The concept of autoencoder was put forward by Rumelhart in 1986. It was used for the dimensionality reduction of complex data, which promoted the development of the deep neural network. Goodfellow et al. put forward that the deep autoencoder can make an exponential decrease in training data and calculation demands. Different denoising network models determine the overall performances of image denoising. It is the core link of study on the whole image denoising process. Therefore, designing the models for autoencoder denoising network is a research priority to scholars. The autoencoder neural network is a translation-based network model, which sets up corresponding relations between the target value and the input value through a back-propagation algorithm. Essentially speaking, image denoising is a translation or conversion process, which converts noisy images into clear images through certain corresponding relationship. This denoising algorithm is non-supervision learning based on the neural network. It displays image characteristics by learning from the hidden layer unit. The output and input can be obtained easily, without considering the changes in the image dimensions. Meanwhile, it can learn good data characteristics. This model generally conducts zero settings for partial nodes during network input, namely the Dropout technology in deep learning. Meanwhile, the algorithm mainly conducts layer-to-layer pre-training through the non-supervision learning of data characteristics, rather than conducts denoising directly. Therefore, the methods improved based on stacked sparse autoencoder has a limited improvement in the denoising effect.

4. Challenges Faced by Deep Learning and Possible Countermeasures

4.1 Means of Model Training
Noisy images can be input into the model and receive training to produce noiseless images. Although it can achieve some denoising performances, the model may get the average noises through fast convergence. It thus restricts the upper limit of denoising performances. Thus the model training can be divided into different batches: A. Noise—noiseless training: Noiseless images are first input to train the model. It aims at enabling the model to grasp good image characteristics and have the capacity of
reconstructing clean images; B. Input noisy images—produce noiseless images. It pays emphasis on testing the model’s noise-detection capacity and abstracting noise characteristics. Combining with the image characteristics learned in previous steps, the model eventually has a stable denoising capacity; C. Adversarial training: Combining with the GAN thinking, the discriminator can be used to offer constant feedback to the effect of image denoising and optimize the network’s denoising performances.

4.2 Limitations of the Model
Although deep learning algorithms are developing rapidly, they generally face tremendous limitations in image denoising. Firstly, deep learning is inapplicable to various noise types and intensities; secondly, many algorithms are only applicable to specific situations; thirdly, robustness is in urgent need of improvement. In consideration of these challenges, more noise types and intensities can be added into the training set. Meanwhile, corresponding sub-models can be pre-trained in different situations. Eventually, all sub-models can be integrated into training to improve the model’s overall de-noising capacity in different noise types and intensities.

5. Conclusions and Prospect
Compared with traditional and classical denoising algorithms, deep learning enjoys continuous development in current adequate computation capacity. Its learning capacity and incomparable development potential are more likely to achieve more satisfactory achievements in image denoising. It is why an increasing number of scholars and researchers are constantly exploring and putting forward different network models and training skills to promote the development of deep learning in image denoising. Deep learning will need a more solid theoretical foundation to develop in the image denoising field. The direction of deepening and expanding the network to get better denoising performances may not be correct. Conversely, researchers may need a deeper examination of the essence behind the issues of image denoising and try to understand the training process of deep learning. In this way, deep learning can be combined with specific problems, rather than simply function as a black box.

References
[1] RadfordA, Metz L, Chintala S. Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks[J]. arXiv preprint arXiv:1511.06434, 2015.
[2] H. C. Burger, C. J. Schuler, S. Harmeling. Image denoising with multi-layer perceptrons, part 1: comparison with existing algorithms and with bounds[J]. Computer Science, 2012, 8-30
[3] V. Jain, H. S. Seung. Natural image denoising with convolutional networks[C]. Neural Information Processing Systems, Curran Associates Inc, 2008, 769-776
[4] J. Allmen, F. Miez, E. McGuinness. Stimulus specific responses from beyond the classical receptive field: neurophysiological mechanisms for local-global comparisons in visual neurons[J]. Annual review of neuroscience, 1985, 8(1): 407–430
[5] Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift[J]. arXiv preprint arXiv:1502.03167, 2015.
[6] J. M. Wolterink, T. Leiner, M. A. Viergever, et al. Generative Adversarial Networks for Noise Reduction in Low-Dose CT[J]. Medical Imaging, IEEE, 2017, 36(12): 2536-2545
[7] C. Szegedy, W. Liu, Y. Jia, et al. Going deeper with convolutions[J]. IEEE Conference on Computer Vision and Pattern Recognition, 2015, 1-9
[8] K. He, X. Zhang, S. Ren. Deep Residual Learning for Image Recognition[J]. 2015, 770-778
[9] Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift[J]. arXiv preprint arXiv:1502.03167, 2015.
[10] I. J. Goodfellow, J. Pougetabdie, M. Mirza, et al. Generative Adversarial Networks[J]. Advances in Neural Information Processing Systems, 2014, 3: 2672-2680
[11] K. G. Lore, A. Akintayo, S. Sarkar. LLNet: A deep autoencoder approach to natural low-light
image enhancement[J]. Pattern Recognition, 2015, 61: 650-662
[12] Arjovsky M, Bottou L. Towards Principled Methods for Training Generative Adversarial Networks[J]. arXiv preprint arXiv:1701.04862, 2017.
[13] Arjovsky M, Chintala S, Bottou L. Wasserstein GAN[J]. arXiv preprint arXiv:1701.07875, 2017.
[14] J. M. Wolterink, T. Leiner, M. A. Viergever, et al. Generative Adversarial Networks for Noise Reduction in Low-Dose CT[J]. Medical Imaging, IEEE, 2017, 36(12): 2536-2545.
[15] N. Divakar, R. V. Babu. Image Denoising via CNNs: An Adversarial Approach[C]. Computer Vision and Pattern Recognition Workshops, IEEE, 2017, 1076-1083.
[16] Y. C. Li; N. F. Xiao; W. L. Ouyang. Improved Boundary Equilibrium Generative Adversarial Networks[J]. IEEE, 2018, 6: 11342-11348.
[17] P. Vincent, H. Larochelle, I. Lajoie, et al. Stacked Denoising Autoencoders: Learning Useful Representations in a Deep Network with a Local Denoising Criterion. [J]. Journal of Machine Learning Research, 2010, 11(12): 3371-3408.
[18] J. Xie, L. Xu, E. Chen. Image denoising and inpainting with deep neural networks[C]. Neural Information Processing Systems, Curran Associates Inc, 2012, 341-349.
[19] F. Agostinelli, M. R. Anderson, H. Lee. Robust Image Denoising with Multi-Column Deep Neural Networks[J]. Advances in Neural Information Processing Systems, 2013, 1493-1501.
[20] H. M. Li. Deep Learning for Image Denoising[J]. International Journal of Signal Processing Image Processing, 2014, 7(3): 171-180.