Research on the Application of Computer Deep Learning in Image Processing

Songchun Zhang¹, Zihao Zhang² and Juliao Qiu³

¹College of Electrical and Information Engineering, Hunan University, Changsha, Hunan, 410082
²Computer Science Department, RWTH Aachen University, Aachen, Nordrhein-Westfalen, 52062
³College of Ocean and Meteorology, Guangdong Ocean University, Zhanjiang, Guangdong, 524088

Abstract. The rapid development of Internet technology has made the whole society enter the era of big data. In recent years, the development trend of artificial intelligence and machine learning has also risen sharply. Informatization has become an important feature of the current era. As an indispensable common information carrier, images not only facilitate people's communication, but also promote the development of deep learning processing image technology. Based on this, this paper analyzes the application of computer deep learning in image processing. Firstly, the deep learning is summarized, its concept and origin are briefly introduced, and then the technical classification, development process and processing purpose of image processing are expounded. Finally, the application of computer deep learning in four aspects is analyzed in detail in image recognition, image denoising, image classification and image enhancement, and it has certain significance for promoting the research and application of deep learning.

1. Introduction
Vision is the main means for humans to obtain a large amount of information in the objective world. Image is one of the manifestations of visual information. With the rapid development of the Internet industry and digital technology, image application has gradually become a part of our daily life, work, and learning [1]. People's single static text has developed into a multi-faceted picture, voice and video with multiple functions [2]. At the same time, the image also provides a platform model for connecting people and services, which has greatly promoted the development of the entire digital life. On this basis, the rapid development of efficient, convenient and accurate target retrieval has liberated a large number of managers and manpower, while meeting the various needs of various users.

2. Deep learning overview
Deep learning is a new research direction in the field of machine learning. It is an important breakthrough in the field of artificial intelligence in the past decade. It has made great success use in many fields such as speech recognition, natural language processing, computer vision, image and video analysis, multimedia, etc. [3]. The foundation of deep learning application is to establish a model to simulate the neural connection structure of the human brain. When processing signals such as
images, sounds and texts, the data features are described by multiple transformation stages, and then given interpretation of data. The concept of deep learning stems from the study of artificial neural networks, which combines low-level features to form more abstract high-level representation attribute categories or features to discover distributed feature representations of data. Existing deep learning models belong to neural networks. Its origins can be traced back to the 1940s, and it was popular in the 1980s and 1990s. The multi-layered perceptron with multiple hidden layers is a deep learning structure. According to the different connections between different layers, the neural network is divided into three types of more important models. There are two kinds of simple applications in image processing. Fully connected to the neural network, that is, the nodes between adjacent layers are strongly connected; the second is the convolutional neural network, the model refers to a convolution kernel, and the adjacent hidden layers are transformed by the convolution kernel, which has become the current research hotspots in the field of speech analysis and image recognition. Its weight-sharing network structure makes it more similar to biological neural networks, reducing the complexity of the network model and reducing the number of weights. The advantage of the network input is multi-dimensional images. The performance is more obvious, so that the image can be directly used as the input of the network, avoiding the complex feature extraction and data reconstruction process in the traditional recognition algorithm. The inter-layer connection and the spatial domain information in the convolutional neural network are closely related, making it suitable for images processing and understanding. Moreover, it also shows superior performance in terms of its automatic extraction of the salient features of the image. In most of the current work, researchers have applied convolutional neural networks to a variety of machine learning problems, including face recognition, document analysis and language detection. No matter which model, in the classification problem, the model optimization algorithm often adopts the form of cross entropy. Although the training method is detailed different, but generally use the reverse training algorithm, although theoretically there are local optimal solutions, but in practice BP method mostly uses random gradient descent, these local gradient information, using local information to calculate the gradient, often can jump out of the local optimal cycle to some extent.

3. Image processing overview

Image processing technology is divided into two categories: analog image processing and digital image processing [4]. Digital image processing techniques are generally processed by computer processing or real-time hardware, and are therefore also referred to as computer image processing. Digital image processing technology converts image signals into digital signals and processes them using a computer or some type of digital processing hardware to improve image usability. Digital image processing first appeared in the 1950s, when electronic computers were developed to a certain level, and people began to use computers to process graphics and image information. Digital image processing was formed as a discipline in the early 1960s. So far, image processing technology has received extensive attention in many application fields and has made significant pioneering achievements, making image processing a new and promising new discipline. In recent years, based on the human visual perception mechanism, an iterative image feature generation method is presented, and it is a research hotspot in image processing. It has great significance and value. Since the development of image processing technology, it is no longer the technology that can only perform simple image processing and recognition, but technology that can infiltrate into various fields and can affect various fields.

In image processing, the input is a low-quality image, and the output is an image with improved quality. Common image processing methods include image enhancement, restoration, encoding, and compression. The first successful application was the American Jet Propulsion Laboratory (JPL). They used image processing techniques such as geometric correction, gradation transformation, noise removal, etc. on the thousands of lunar photos sent back by the Space Detector No. 7 in 1964, taking into account the position of the sun and the environment of the moon. The impact of the successful mapping of the moon's surface map by the computer has been a huge success. Later, more complex
image processing was performed on the nearly 100,000 photos sent back by the spacecraft, so that the topographic map, color map and panoramic mosaic of the moon were obtained, which achieved extraordinary results and laid a solid foundation for human landing on the moon. The foundation also promoted the birth of the discipline of digital image processing. In the future research of inter-aircraft technology, such as the exploration of planets such as Mars and Saturn, digital image processing technology has played a huge role. Another great achievement in digital image processing is the result in medicine. In 1972, the British EMI company engineer Housfield invented the X-ray computed tomography device for head diagnosis, which is what we usually call CT (Computer Topography). The basic method of CT is to reconstruct a cross-sectional image by computer processing according to the projection of the human head section, which is called image reconstruction. In 1975, EMI successfully developed a CT device for the whole body, which obtained a clear tomographic image of various parts of the human body. In 1979, this non-invasive diagnostic technology won the Nobel Prize, indicating that it has made an epoch-making contribution to humans. With the in-depth development of image processing technology, from the mid-1970s, along with computer technology and artificial intelligence, thinking science with the rapid development of research, digital image processing has developed to a higher and deeper level. People have begun to study how to interpret images with computer systems to achieve an understanding of the external world similar to the human visual system. This is called image understanding or computer vision. Many countries, especially developed countries, have invested more manpower and material resources in this research and have achieved many important research results. The representative result is the visual computing theory proposed by Marr of MIT in the late 1970s. This theory has become the dominant idea in the field of computer vision for more than a decade. Although image understanding has made great progress in the study of theoretical methods, it is a difficult research field itself. Because human beings still have little understanding of their own visual processes, computer vision is awaiting new areas of people's further exploration.

The purpose of image processing:

1) Improve the visual sensory quality of the image, producing images that are more suitable for human visual observation and recognition. Such as image brightness, color conversion, enhancement, image geometric transformation, etc., to improve the quality of the image.

2) Extracting certain features or special information contained in an image facilitates computer analysis of images. The extracted features can include many aspects such as gray/color characteristics, boundary/region characteristics, texture characteristics, shape/topology characteristics, and relational structures.

3) The image data is transformed, encoded and compressed to improve information transmission efficiency and reduce image information storage capacity. Image and video information is large and often requires effective compression of such data.

4) Using techniques and methods in graphic images to help people understand and analyze data, that is, information visualization.

5) The scope of information security is very wide, ranging from national military politics to the disclosure of personal information. It can be monitored and identified using digital image processing techniques.

4. Application of deep learning in image processing
Since all image operations inevitably distort some inherent relationships between adjacent pixels in the image. Therefore, many scholars have conducted in-depth research on the inconsistency of the inherent relationships between these adjacent pixels. Interestingly, the original intention of deep learning research is mainly applied to image recognition. So far, although deep learning has been applied to speech, images, text, etc., about 70% of the papers published in the field of deep learning are about image recognition [5]. From the ImageNet competition in 2012, deep learning has played a huge role in the field of image recognition. In the fields of general image classification, image detection, optical character recognition, face recognition, etc., the best systems are based on deep learning. The three advantages of deep learning are reflected in the recent progress of image
recognition: the model structure is more and more complex, and the scale of training data is increasing; various prior knowledge about data structure is reflected in the new model structure; End-to-end learning has made us increasingly discard intermediate steps based on manual rules.

Traditional image processing tasks rely on artificially designed features for a long time, and the characteristics of artificial design are often not comprehensive and in-depth. They can only extract low-level information of images, but it is difficult to systematically extract deep-level information. Deep learning can automatically identify more advanced features in the image. This feature subverts the traditional image processing method and achieves an unprecedented recognition effect [6]. When processing image data, the spatial information is often extremely valuable, which has a great influence on the final model effect. For example, in the face recognition, the position of the eye relative to the nose is recognized, and the recognition of the eye itself is more accurate. Feedforward neural networks and cyclic neural networks cannot capture spatial information in images, and convolutional neural networks can perfectly handle this problem, and can quantify high-level spatial topology information in images to achieve adaptive feature recognition. The effect is much stronger than the traditional manual feature extraction.

4.1. Application of Deep Learning in Image Recognition

Image recognition belongs to the category of pattern recognition. The main content is that after some pre-processing of the image, image segmentation and feature extraction are performed to judge the classification. Statistical pattern classification and structural pattern classification are commonly used pattern recognition methods. The deep trust network and the stack self-encoding network perform well in a single image recognition task, successfully used to generate compact and meaningful image retrieval representations, and have been used in large image retrieval tasks with very good results [7]. The conditional deep trust network is used for video sequencing and human motion synthesis. The conditional deep trust network makes the weight of the deep trust network associated with the previous data, which can improve the effectiveness of the training. Lee and Raina et al. used sparse coding and deep trust networks to learn effective feature representations from natural images. Nair et al. proposed an improved deep trust network. The top-level model of the model uses a third-order BM. They use this model for the three-dimensional target recognition task NORB data set. The experimental results show that the training has a low prediction error rate. Tang et al. proposed two strategies to improve the robustness of the deep trust network. First, the first layer of the deep trust network is introduced into the regularization method with sparse connection structure, and then a probabilistic noise reduction algorithm is proposed. These techniques are in high noise images. The robustness of the recognition task and random noise shows its effectiveness. Lee et al. proposed a deep learning method to automate brain image segmentation, and use convolutional neural networks to establish discriminant features for brain image segmentation. It can automatically learn from the class labels provided by human experts, and verify that the method is automatic. The multi-class brain image segmentation shows superior performance, which indicates that this method can replace the existing template image segmentation method, which reduces the intervention of human image experts and the need for prior information in the image segmentation process.

In recent years, the development of deep learning in image recognition has the following main trends:

1) Model level is deepening

The network used in the 2012 ImageNet competition champion is five convolution layers (including three pool layers and two norm layers). By 2014, the champion used 59 convolutional layers (including 16 pool layers and 2 norm layers). The second place also used 19 convolution layers and achieved better performance. The importance of the depth of the model is self-evident.

2) Model structure is becoming more complex

Traditional convolutional neural network models are mostly built using simple conv-pool-norm structures. The results show that the parallel multi-resolution inception structure can fuse the effective information of images at different scales, while the structure NIN (network-in-network) compresses
the convolutional layer of larger parameter scale by low rank decomposition, which greatly reduces the model parameter size. In doing so, on the one hand, it can reduce the degree of over-fitting and improve the promotion ability of the model, on the other hand, it provides very favorable conditions for large-scale parallel training.

3) Massive annotation data and appropriate data perturbations

Combined with the characteristics of image data, data perturbation methods including translation, horizontal flipping, rotation, scaling, etc. They are used to generate more effective training data, which can generally improve the promotion ability of the recognition model.

4.2. Application of Deep Learning in Image Denoising

Image denoising is a very important part of image processing. At present, deep learning has been widely used in the field of image denoising algorithms. The denoising method can effectively improve the computer's ability to recognize image information, and its application in security and medical fields is very important [8]. It is not difficult to use the convolutional neural network learning model to learn image features. In recent years, the application of image denoising has been advancing very fast. For example, the image denoising algorithm based on the deep learning model is widely used. By establishing a nonlinear mapping between the original image and the noise image, the convolution subnet is used to collect the image feature information, and then image feature information is used to restore the original image. In actual use, the denoising method of such deep learning can extract a large amount of texture information in an image, and then has a very good effect on image processing.

4.3. Application of Deep Learning in Image Classification

The most important advancement in deep learning in object recognition is reflected in the image classification task in the ImageNet ILS-VRC3 challenge [9]. The lowest error rate for traditional computer vision methods on this test set is 26.172%. In 2012, Hinton's research team used a convolutional network to reduce the error rate to 15.315%. This network structure is called Alex Net. Compared with the traditional convolutional network, it has three differences: First, Alex Net adopts the dropout training strategy, which will input the input layer and the middle layer during the training process. The neurons are randomly set to zero. This simulates the various disturbances of noise to the input data that cause some neurons to miss detection of some visual modes. Dropout converges the training process more slowly, but the resulting network model is more robust. Second, Alex Net uses a rectified linear unit as a nonlinear excitation function. This not only greatly reduces the computational complexity, but also makes the output of the neuron sparse and more robust to various disturbances. Third, Alex Net generates more training samples by mirroring the training sample and adding random translation perturbations, reducing overfitting.

The classification algorithms involved in image processing mainly include three steps: region division, feature extraction and classifier recognition classification. The most important step in the three steps is the extraction of image features, whether the image can be extracted accurately. It will directly affect the subsequent classification and identification work. Deep learning, because of its many advantages in the field of computer vision, is also widely used in image classification algorithms. Reasonable application can improve the feature extraction ability. For example, a deep learning-based image classification calculation method, which is widely used, uses a single-marked image and deep learning between the labeled images, and the single-marked image is characterized by an LDA algorithm and a PCA algorithm. The dimensionality reduction process is performed, and then the classification operation is performed by the KNN analyzer and the SVM analyzer to realize the dimensionality reduction processing of the image, which can then optimize the image classification [10]. At the same time, the image features in the more complex image classifications such as multi-label images can be extracted by comparing the average hausdorff and the minimum hausdorff distance metrics.

With the continuous development of artificial intelligence technology in China, the application of image classification feature extraction based on deep neural network is deepening and the practicality
is also increasing. Taking Baidu in China as an example, Baidu search engine based on deep learning can achieve more than 90% accuracy in image classification and recognition. In addition, the accuracy of Alibaba's Alipay scanning image recognition in image classification is also 90% or so. However, the traditional image classification feature recognition is difficult to break through 50% in accuracy, which shows the great advantage of deep learning in image classification applications. It can be expected that with the continuous maturity of big data technology and artificial intelligence technology, the accuracy of image classification based on deep learning is expected to reach 100%.

4.4. Application of Deep Learning in Image Enhancement

Image enhancement is to improve the quality of the image, improve the sharpness of the image, and so on. It highlights a portion of an image in accordance with specific requirements while attenuating or removing certain unwanted information processing methods. Its main purpose is to make the processed image more suitable for an application. Histogram modification processing, image smoothing processing, image sharpening processing, and color processing techniques are currently methods of image enhancement. Image enhancement is an important and indispensable part of image processing. Reasonable image enhancement can effectively enhance the feature area of the image, and then further improve the visual effect of the image. Subsequent machine and manual identification have significant advantages. Deep learning is also widely used in image enhancement algorithms. Image enhancement is improved by super-resolution algorithm in convolutional neural network. White balance can even correct the color of underwater images correctly. There is also a clear advantage in the resolution of the image. It can be seen that the image based on the deep learning enhancement process can meet the higher visual requirements of the image.

As an essential stage of image processing, the result of image enhancement can highlight the feature areas in the image and improve the visual effect of the image, so that the enhanced image can better identify humans and machines. Image super-resolution technology combined with deep learning theory for image enhancement processing, the super-resolution algorithm of convolutional neural network and fast convolutional neural network is improved, so that the processed image more satisfies people's visual requirements. Experiments show that the proposed algorithm effectively increases the effect of image reconstruction and provides a certain reference value for future image enhancement techniques. The convolutional neural network is used to enhance the image, and the white balance method is used to effectively improve the color correction of the underwater image. The scene depth model is used to deblur the operation, and the super-resolution model refines the resolution of the image. It is proved that this technology has certain improvement in the research of underwater image recognition resolution.

5. Future development prospect

The application of deep learning in image recognition is in the ascendant, and there is a huge development space in the future.

One of the trends in object recognition and object detection is to use larger and deeper network structure. In ILSVRC 2012, Alex net only contains five volume layers and two full connection layers. In ILS-VRC2014, gooleNet and VGG use more than 20 layers of network structure. The deeper network structure makes back propagation more difficult. At the same time, the scale of training data is also growing rapidly. It is urgent to study new algorithms and develop new parallel computing systems to train bigger and deeper models with big data more effectively.

Compared with image recognition, the application of deep learning in video classification is far from mature. The image features from Imagenet training can be directly and effectively applied to various image related recognition tasks (such as image classification, image retrieval, and object detection and image segmentation) and other different image test sets, with good generalization performance. However, deep learning has not yet obtained similar features for video analysis. To achieve this goal, not only large-scale training datasets should be built, but also new depth models suitable for video analysis should be studied. The amount of computation for training the depth model
for video analysis will also be greatly increased. In applications related to image and video, the output prediction of depth model (such as segmentation image or object detection frame) often has spatial and temporal correlation. Therefore, it is also a key point to study the depth model with structural output.

6. Conclusion
In the past few years, thanks to deep learning algorithms, the research and application of image recognition technology has developed rapidly. Image annotation, object detection, object segmentation, pose estimation, face recognition, optical character recognition, almost all classic image recognition techniques have made breakthroughs with the help of deep learning algorithms. Throughout the current research status in the field of imagery, deep learning leads almost all frontier development directions, but there are still some other limitations. For example, deep learning can be compared with traditional machine learning algorithms under the support of computing resources and data volume. It shows more excellent effects, but the requirements and consumption of computer performance are even greater. At present, the research on neural network mostly exists in the design of some new architectures, and there is no new energy injection for deeper training methods. The development of the future image processing field will undoubtedly accompany the pace of deep learning. It is expected that deep learning will make significant progress in the basic theoretical fields such as models and training methods, so that it can be better applied and practiced.

References
[1] Nan Zhe, Qian Yurong, Xing Yanni, Zhao Jingxia. A Survey of single image super resolution reconstruction based on deep learning [J]. Computer Application Research. 2018 (6): 265-268.
[2] Sun Xu, Li Xiaoguang, Li Jiafeng, Zhuo Li. Research progress on image super-resolution restoration based on deep learning [J]. Chinese Journal of Automation. 2017 (05): 127-129.
[3] Hu Changsheng, Zhan Wei, Wu Congzhong. Image super-resolution reconstruction based on depth feature learning [J]. Chinese Journal of Automation. 2017 (05): 202-207.
[4] Zhu Xinshan, Qian Yongjun, Sun Wei, Ren Chao, Sun Ya, Yao Siru. Image restoration forensics algorithm based on deep neural network [J]. Journal of Optics. 2018 (11): 78-82.
[5] Yang Bin, Zhang Tao, Chen Xianyi. Image local fuzzy recognition based on deep learning [J]. Journal of Applied Science. 2018 (02): 191-194.
[6] Mark Everingham, Luc Gool, Christopher K. I. Williams, John Winn, Andrew Zisserman. The Pascal Visual Object Classes (VOC) Challenge [J]. International Journal of Computer Vision. 2010 (2): 232-235.
[7] Karen Simonyan, Andrew Zisserman. Very deep convolutional networks for large-scale image recognition [J]. Corrosion. 2014 (08): 350-356.
[8] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition [J]. IEEE Conference on Computer Vision and Pattern Recognition. 2016 (12): 1012-1017.
[9] Chen J, Kang X, Liu Y, et al. Median filtering forensics based on convolutional neural networks [J]. IEEE Signal Processing Letters. 2015 (08): 783-788.
[10] Jiangqun Ni, Jian Ye, Yang YI. Deep learning hierarchical representations for image steganalysis [J]. IEEE Transactions on Information Forensics and Security. 2017 (10): 1120-1126.