

Research Article

Biological Image Processing Algorithm Based on Attention Mechanism and Convolutional Neural Network

Lisong Ou 

College of Science, Guilin University of Technology, Guilin 541000, GuangXi, China

Correspondence should be addressed to Lisong Ou; oulisong@glut.edu.cn

Received 22 September 2022; Revised 20 October 2022; Accepted 25 November 2022; Published 15 April 2023

Academic Editor: Nagarajan Govindan

Copyright © 2023 Lisong Ou. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper proposes a model algorithm based on convolutional neural network combined with attention mechanism to realize fast and accurate identification of biological image. Firstly, deformable convolution is used to extract features in the horizontal and vertical directions, respectively. Secondly, attention modules are used to capture remote dependencies in one spatial direction, while accurate position information is retained in another spatial direction, so that information in both vertical and horizontal directions can be retained; after a series of transformations, the attention vector is obtained and multiplied back to the original feature vector as a weight factor. The experimental results show that the proposed algorithm can effectively improve the image quality, improve the image clarity, avoid color distortion, and achieve good results in both synthetic and real low-illumination images, and the subjective and objective evaluation indicators are better than the contrast algorithm.

1. Introduction

In recent years, convolutional neural network has been introduced into various applications in the biomedical image field, such as disease classification, prediction, focus detection, and image registration, and has achieved great success. Compared with traditional machine learning methods, the deep learning method based on the convolutional neural network can eliminate the complexity of feature engineering and complete the task end-to-end through the strong feature expression ability of the convolutional neural network [1]. Since the end of 1990s, the recognition technology of the convolutional neural network has entered a rapid development process and gradually maintained stability. However, the application of the convolutional neural network in the field of medical images still faces many challenges, including the small amount of medical image data leading to inadequate model training and poor performance; some imaging instruments have poor imaging quality, which increases the difficulty of model optimization [2–4]. This paper analyzes the existing convolutional neural network structure in detail [5] and then proposes a convolutional neural network using attention mechanism to improve network performance.

2. Overview of Research

2.1. Research Background. Since the advent of 5G technology, due to the rapid development of social economy, the upgrading rate of computer information technology has become faster and faster, and it has penetrated into all walks of life, bringing great convenience to people. With the popularity of the Internet, the amount of information has exploded, and the amount of information has greatly increased [6, 7]. Therefore, the convolutional neural network technology can improve the accuracy of image and more effectively explore the information of image data. The application of artificial intelligence is becoming more and more perfect, and the deep neural network has attracted everyone's attention. The traditional image recognition system is relatively outdated, and its accuracy is low [8, 9]. The research shows that the traditional identification method cannot meet the current practical needs for a large amount of image information.

In 1998, LeCun published a paper proposing LeNet-5, which is the origin of the convolutional neural network. Then, in 2012, in the ImageNet image recognition contest, the Hinton group proposed AlexNet, which increased the

top-5 error rate of the contest results from 26% to 15%, and overturned the image recognition field. As a result, the CNN began to rise. After years of development, its performance has become stronger and stronger, and it has excellent performance in a variety of computer vision tasks. It has also been applied to the field of medical images and has become a research hotspot today. There are many research studies devoted to improving the performance of the CNN. In terms of the depth and width of the network, attention mechanism is also a way to enhance the ability of the CNN by focusing on effective features and suppressing unimportant features. Attention mechanism is also used in the field of medical images. Zhang et al. used attention residual network to classify skin injuries.

2.2. Research Purpose. In recent years, the deep learning optimization algorithm of the convolutional neural network has developed rapidly. At the same time, deep learning and convolutional neural networks are widely used in speech recognition technology, target detection, image classification, and other fields [10] and are constantly improving and progressing. On the other hand, image classification belongs to the category of image classification. Scientific research on how to identify and classify images quickly and clearly through deep learning and convolutional neural networks is of great significance to the social and economic development, which will help everyone understand image classification. Image classification is a research subject in the field of machine vision and image resolution, using the image information contained in the image to distinguish different image types.

The relative technical component of artificial intelligence is a powerful weapon to build a technical postfortress. Product development does not require the product manager to type the code himself. After mastering the technical principles, such professionals can turn business processes into product problems that can be implemented quickly and reasonably, which is crucial for the research and development of AI-related products, and they also need to have a certain understanding of the current image processing technology [11, 12].

From another point of view, the convolutional neural network (CNN) is essentially a universal artificial intelligence technology for image processing, and it is an algorithm technology with feedforward neural network. In 2012, Alex Krizhevsky won the ImageNet competition with the convolutional neural network technology. After that, Alex reduced the error record of biological image classification from 26% to 15%. Facebook, Google, Amazon, and other well-known companies all use the convolutional network for biometric image recognition and product push. Generally speaking, common neural networks are completely connected to the input layer and the hidden layer to ensure that the system can extract biological image features. From the point of view of computing power, it is feasible to compute features from the whole biological image. In contrast, the calculation process of $96 * 96$ biometric

images is 100 times slower than that of $28 * 28$ biometric images. As we all know, nowadays photos are always high-definition biological images [13], and ordinary neural networks cannot predict when they will be completely processed properly.

3. Research Status at Home and Abroad

3.1. Technical Status. Convolution is the process of moving an image by using a dialog box with a fixed size, and then multiplying the convolution dialog box by the definition to obtain the exported image. The convolutional neural network (CNN) is complex in explaining the calculation process, but it needs a deep understanding of the CNN optimization algorithm [14]. After nearly 30 years' development, the convolutional neural network has become a branch of the Internet and developed rapidly, promoting the operation of VGG16 and VGG19 in the transport layer, so as to strengthen the development of technologies such as the convolution control module NIN Internet and R-CNN. It won many new networks from category to target detection.

3.2. Application Scope and Field. In the application of image classification, it is necessary to analyze the feature expression of images. Especially in the method of image classification based on machine learning, image feature expression is very important as the training and prediction analysis of classification model. Before the in-depth analysis of image feature expression, human-defined image features are the most widely used feature expression methods. According to the image category described by features, the features defined by manpower can be divided into global features and local features.

The overall feature key describes the overall attributes of an image, which is a pixel-level, low-level, and visual effect feature. The color, texture, and appearance of an image are general global features. The key of color tone is to analyze features according to other statistics such as bar graph, cumulative bar graph, mean gray value, mean value, and standard deviation. It involves simple calculation and has stable quality. The texture has the ability to mainly show images with large differences in thickness and relative density. The characteristics of traditional global texture include wavelet analysis and gray level co-occurrence matrix. Compared with color and texture features, appearance features have word meaning information, so the contour and regional features of the overall target object in the image can be obtained as much as possible. However, the expression of appearance features is generally image segmentation, and many existing image segmentation algorithms are not completely accurate, so it is difficult to obtain appearance features. The existing appearance attributes of include binarization threshold, Canny boundary definition, nearby definition, milky white area, and white area. The overall feature can indicate the overall information of the image, but it is difficult to describe the details of the image, and the utilization rate of image information is weak.

Local feature key describes the local area of an image and attaches importance to the attributes of image details, such as edges, corners, and curves. Classical local features include local binarization (LBP), Gabor wavelet transform (SIFT), and accelerated robustness feature (SURF). Many effective image processing methods are given. In the other way, the Hal local binary solid model feature (HLBP) is used to express the local texture features of the image, which can effectively obtain the texture information of the deep image with low computational complexity. Sun Jun et al. gave a multiscale texture identification method based on LBP and direct evidence theory of damping filtration. Experiments show that this method has advantages.

Global and local image features describe different image attributes, but the image attributes cannot be completely described according to the image features of a single mode. Therefore, some scholars fuse different types of features to fill the advantages of features and image features.

3.3. Research Status at Home and abroad. The convolutional neural network (CNN) is outstanding in many industries of natural image understanding. With the in-depth analysis of theory and application, the application of the convolutional neural network in medical image resolution has attracted the attention of scholars. The convolutional neural network can learn the essential characteristics of images. After nonlinear mapping, it is mainly used for classification and reconstruction to solve different daily medical tasks. Of course, CNN's excellent image understanding results can be used as a reference for daily tasks of different medical clinical research such as image deconvolution of fluorescent microscope, cone beam, iterative updating and reconstruction of statistical analysis of CT (CBCT), and PET image segmentation. Image deconvolution is an image processing method to reduce image blur, noise level, and image uncertainty. According to the operation process of detecting the decline of the training image, the best possibility can be obtained according to the original CNN image. Compared with the classical deconvolution method, this method not only reduces the image noise, suppresses the image artifacts, and retains the details of the image but also has strong generalization ability, and can use specific numbers well. CBCT reconstruction technology is a technology that uses X-ray projection data information to reconstruct the cross-sectional image of the patient's body from the body, thus assisting doctors in clinical diagnosis and treatment.

According to CBCT of the CNN model, the statistical analysis iterative updating reconstruction algorithm is improved, which reduces the harm of Hessian penalty to image ambiguity and further improves the quality of image reconstruction. This optimization algorithm can not only suppress the noise in the image and clear the steps but also improve the screen resolution of the image. Image segmentation of PET plays an important role in clinic, which is dedicated to defining the size and location of diseases and assisting clinical medical diagnosis and treatment. Softmax classifier based on the training texture features can predict and analyze the types of sharpness well, but it is also very

prone to misclassification. According to the multifeature segmentation method of PET images, the CNN has both the texture features and the initial gray value features of PET images, and the experimental conclusion has good accuracy and scalability. Generally speaking, the convolution neural network depth model can promote the development of medical image processing and various medical image solving daily tasks, which is of great significance to clinical medical diagnosis and treatment.

The practical operations related to the input of image enhancement function formulas around the world include arbitrary shearing, arbitrary inversion technology, arbitrary saturation adjustment, arbitrary screen brightness level, standardization, and integration to form more training samples. In many image recognition problems, the pre-processing of image data can avoid the influence of irrelevant element model fitting and can also achieve the purpose of improving the data and the accuracy of the model.

The convolutional neural network (CNN) has achieved unprecedented success in the computer vision industry. At present, many traditional computer vision optimization algorithms have been replaced by deep neural networks. Because of its huge economic benefits, the deep neural network and convolutional neural network have become research hotspots, and many classic works are varied. Image identification is the core and basic problem of computer vision. Other daily tasks, such as target detection, image segmentation, image formation, and video understanding, are relatively highly dependent on the linguistic expression ability of features in image identification. The latest news of image recognition directly affects the performance of all computer vision daily tasks based on deep learning, so it is particularly important to fully understand this progress.

4. Convolution Neural Network and Attention Mechanism

4.1. The Structure and Characteristics of the Convolutional Neural Network. The convolutional neural network (CNN) is very similar to the ordinary neural network. This is composed of neurons with learning and training weights and deviations. Each neuron receives some inputs and performs some dot product calculations by export type. This is also applicable to some high-performance calculations in general neural networks. The default setting input of the convolutional network is the image, which can number the special characteristics in the network architecture, making the feedforward control more efficient and reducing many main parameters. Neurons have a three-dimensional volume, which is designed to be three-dimensional in total width, extreme and deep according to convolutional neural networks. It should be noted here that these deep layers are not the depth of neuron network, but only used to describe neurons. For example, if the input image size is $32 * 32 * 3$ (RGB), the input neurons also have three-dimensional values of $32 * 32 * 3$. The actual explanation is that the convolutional neural network is composed of several layers, with three-dimensional input and three-dimensional

output. Some layers have main parameters, while others do not. Convolutional neural networks generally include the following layers.

The first layer is the convolution layer. Each convolution layer of the convolutional neural network consists of several convolution sum modules, and the main parameters of each convolution sum unit are optimized according to the back propagation algorithm. The purpose of discrete convolution is to obtain different features of the input. The first convolution layer can only obtain low-level features such as edges, parallel lines, and corners. Double-layer Internet can iteratively update low-level features to obtain more complex features. The second layer is the linear rectifier layer (relu layer). The nerve excitation function of the system is linear rectifier (relu). Generally, the pool layer behind the convolution layer will get a larger size primitive. Cut the element into several regions, and take the maximum or average value to obtain a smaller new element. All partial features are classified into global features in the whole connection layer, and the final score of each type is estimated. The third layer is the full connection layer, which fuses all partial features into global features and calculates the final score of each type.

4.2. Attention Mechanism. The basic idea of attention mechanism is to focus on the key information that is important to the target task, while ignoring the unimportant information, which is borrowed from the human visual attention mechanism. The realization of this kind of attention is to obtain the weight through learning. The weight range is $[0, 1]$, which represents the importance of the feature, and then weight the input to complete the attention process. Attention mechanism has been extensively studied and applied in machine translation and natural language processing. In the field of image analysis, there are also some studies. Hu et al. introduced a compression excitation module that uses global average pooling to calculate channel attention. Woo et al. proposed a CBAM module that acts as the attention module of the convolutional neural network to generate attention in both space and channel.

4.3. Biological Image Recognition and Convolution Neural Network. Defining the features of human images can describe the structure, texture, and other characteristics of images, and it is the key to reflect the basic meaning of images. For images with simple structure, manually defining features can better describe features. In complicated structural images such as slate, the features defined by manpower are not very appealing. The convolutional neural network can effectively acquire the word meaning of the upper layer of the image, which is mainly used for the automatic features of the image. Through continuous trials and tests, the researchers first gave the automatic learning image features of the supervised back propagation network LeNet, and completed the data image identification. The LeNet network can be regarded as the prototype of the convolution neural network. The LeNet network model framework is shown in Figure 1.

In recent years, VGG, GoogleNet, and ResNet have a deep network hierarchy and strong feature expression ability. The convolutional neural network has already caused extensive attention and scientific research. Many researchers have improved the existing convolutional neural network and applied it to different image features. According to the word meaning characteristics of the pen-written images obtained by the LeNet-5 network, the hidden Markov model (HMM) is fused to identify the words composed of Saudi Arabia identifiers. Bai Cong et al. gave a deep learning framework, which can be used for large-scale image classification. It can enhance and improve the network architecture and internal structure of AlexNet, so that the network can express image features.

With the increase of network level, the total number of main parameters of the network is greatly increased, and the number of logo samples in the network is also greatly increased. However, because of the high cost of data marking, many feasible APP applications may be difficult to meet this requirement. At present, some researchers have focused on the scientific research of image features of small-scale data sets.

In order to flexibly use the limited logo template, the image is prepared by a series of arbitrary transformations. Common methods include image resolution rotation, scaling, reflecting surface, rotation, cropping projection, and scaling. Some researchers take advantage of the reusability of deep neural networks, such as VGG and Google Search Network. They practice on big data sets such as ImageNet, and feature and classify images with small data sets. For the small-scale data set close to the initial data set, we can also use the convolution layer and all-connected layer of the pretraining model as image features, fuse svm algorithm and other classification models to complete the classification, and obtain the convolution neural network export of different levels trained by Krizhevsky and others as image features, thus completing daily tasks related to machine vision such as target detection and scene recognition. According to the pretrained deep convolution entity model, other researchers adjusted the hazards of new data sets, image features, and classification, and adjusted convolution and network on smaller data sets according to the layer freezing method to complete image classification.

In addition, the number of logo templates can be reduced by reducing the main network parameters, such as adjusting the number of network overlay layers and the size of each network layer, and standardizing the main parameters to prevent overfitting. The meaning of the convolution layer is to form multiple images from an original image according to convolution calculation, and maintain the initial features of each image. This convolution calculation effect is very similar to the principle of filter lens, and it is used to obtain different types of biological image features. In the process, the input pixel value is $6 * 6$, and the convolution core pixel value is $3 * 3$. At first, a $3 * 3$ matrix (blue block) of pixel values at the top left of the image is selected for convolution kernel calculation. Its internal process is as follows.

The product of the selected area and the primitive matched with the convolution kernel position is a $3 * 3$ pixel

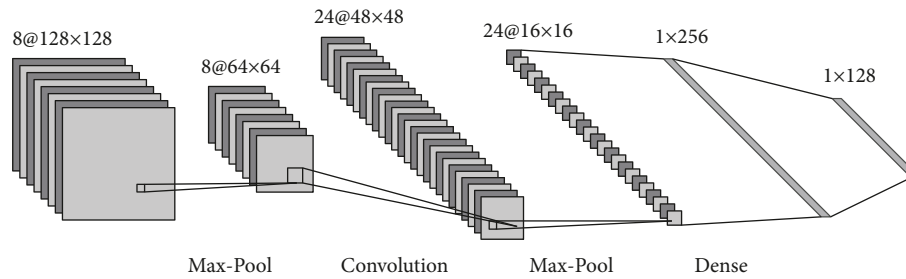


FIGURE 1: LeNet network model framework.

matrix, and all the pixel values included in the matrix are summed to obtain the original convolution features as follows:

$$3 \times 1 + 0 \times 0 + 1 \times (-1) + 1 \times 1 + 5 \times 0 + 8 \times (-1) + 2 \times 1 + 7 \times 0 + 2 \times (-1) = -5. \quad (1)$$

Then, move the dark blue selected area to the right one step, and use the newly obtained matrix to calculate the convolution of the convolution kernel. You can also carry out multiplication first and then add. The convolution feature at the corresponding position is -4 . Then, the dark blue ps selection area is arranged from left to right and from top to bottom. When all the images are calculated, the eigenvalues of the images obtained after the convolution operation are all matrices with pixel values of $4 * 4$. In the experimental step, different types of convolution verification images are usually used for inpainting, so as to obtain different types of convolution numerical features.

From the whole convolution process mentioned above, it is known that according to convolution calculation, the image size decreases from the initial value of $6 * 6$ to the numerical value of $4 * 4$. If the image is convolved several times, the edge features of the image will often be ignored. In order to maintain the stable size of the image and preserve more image features in the convolution process, it is necessary to get rid of the image and fill in the blank area outside the image, usually with a filling value of 0. The basic deep convolution network structure is shown in Figure 2.

5. Convolutional Neural Network for Biological Image Recognition

5.1. Mapreduce Algorithm Scheme. Mapreduce is a parallel processing programming entity model, which is mainly used in the work of large- and medium-sized computer clusters, and can effectively solve TB and PB data sets. The programming entity model has the advantages and characteristics of strong inclusiveness, easy use, and overall expansibility, and has been widely used in the computer industry. Mapreduce programming structure is a tree structure of large- and medium-sized data sets using the master node management method. The master node will allocate resources to different single connected areas, and solve and converge the data from the child nodes to the

master node. The whole process of Mapreduce is mainly composed of Map function and reduce function, and the specific solutions to daily tasks of the two functions are quite different. Map function decomposes the work plan into several control modules, and Rudeuce summarizes the functions of the modules.

The algorithm divides the data after practice into basically the same small parts, and effectively stores the corresponding nodes on Hadoop website to make them evenly distributed. During the training period, the data from the CNN network is stored in another node. Mapper's daily task accepts data that can be applied to different nodes, and calculates the change of weight and offset by forward and fixed term transmission to generate positive and intermediate keys. After calculation, it solves and summarizes the local documents.

5.2. CUDA Algorithm Scheme. The CUDA programming model can be applied to the processing and calculation of GPU data development and use environment. It was put forward by NVIDIA enterprises. This is also an image processing device. The use of GPU has greatly improved the efficiency and quality of computer graphics, and promoted the development of computer image simulation, virtual reality technology environment, and image processing technology. CUDA and GPU Cheng Xuyuan can effectively integrate CUDA and C language selection and editing service platforms, and use GPU to carry out the whole execution process of the source program, without having to independently grasp the professional knowledge of images. Furthermore, the difficulty of GPU data information calculation is reasonably solved, the data information calculation is simplified, and the stability of the whole system is greatly improved. The structure of data processing method on CDA platform consists of CPU and GPU. GPU architecture can turn on computer features and save resources to the greatest extent. In the process of image identification, in order to keep practicing the process and speed, it is necessary to prepare the image according to the situation, reasonably

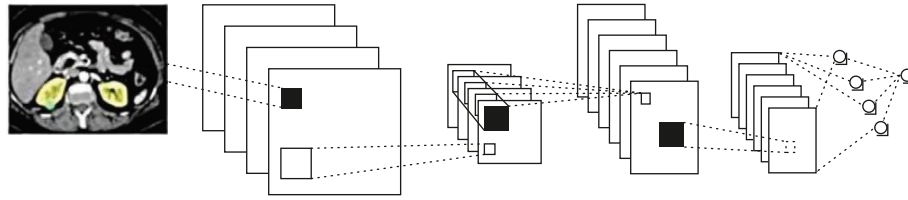


FIGURE 2: Basic deep convolution network structure.

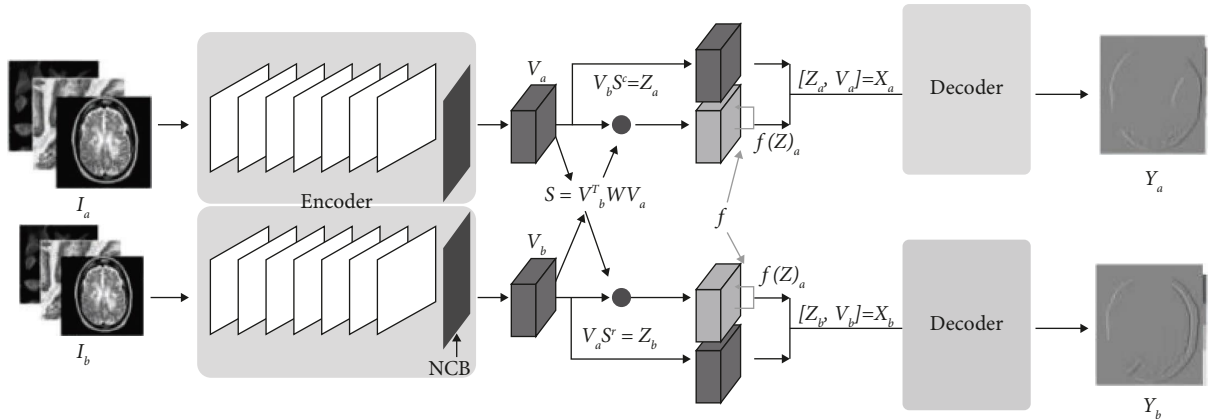


FIGURE 3: Multiconvolution kernel calculation process.

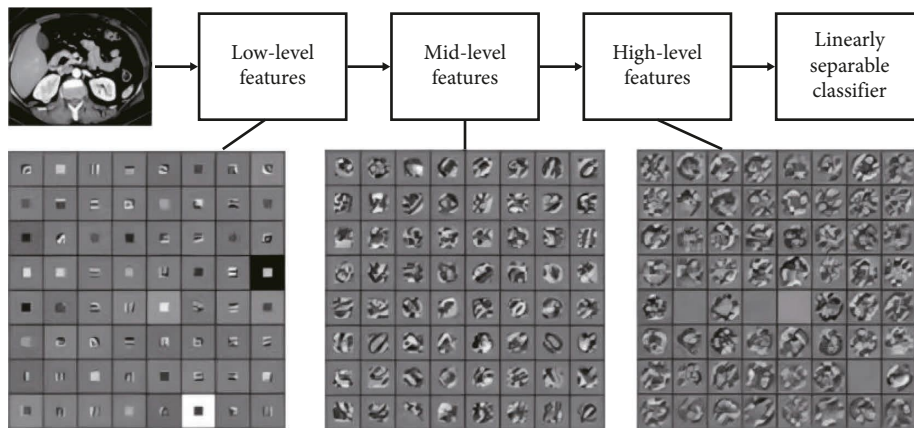


FIGURE 4: Analysis of biological image recognition results.

remove the excessive noise data information in the image, and then segment and identify the processed image.

5.3. Biometric Image Processing Based on Cooperative Attention Mechanism. Due to the problems of low signal-to-noise ratio, target blur, and sound shadow in medical images, traditional methods cannot accurately locate and segment the target, and have the defects of not accurately placing the focus on the target area and not distinguishing the target from the background boundary, which makes it difficult for observers to accurately locate the target position. To solve the above problems, we propose a biological image segmentation method based on cooperative attention. It has two inputs. After feature encoding, cooperative attention process, and feature decoding, the segmented output is

obtained. The biological image based on collaborative attention is shown in Figure 3.

5.4. Analysis of Biological Image Recognition Results. In the whole algorithm recognition process, the data entered into the system are mainly divided into two directions. When the predictive analysis classification is consistent with the expectation, the system shows the correct classification; when the predictive analysis classification is inconsistent, the system shows the wrong classification.

The COFAR test data are mainly used here. The system effectively stores each actual test result and error rate, and Python is used to check the overall change of error rate in scripting. There are 128 samples in each issue. The classification error rate of the system is 18% without adjusting the

logistic regression calculation of the main parameters and the main parameters of the network architecture. The systematic classification error rate is 13-14%, which can achieve this accuracy, and it is the highest level in history. Not all images are cropped in the initial image solution process, but the images are cropped in the application process. This method takes less time in training, but in the testing session, it takes a significant increase. The analysis of biological image recognition results is shown in Figure 4.

6. Conclusion

In this work, we presented the biological image recognition application based on the convolutional neural network and attention mechanism. Among them, the convolutional neural network has great advantages in image analysis, and has been applied in disease classification, prediction, and image registration, and has achieved great success. Due to its traits of focusing on key statistics and suppressing unimportant information, attention mechanism can radically enhance the overall performance of the enhanced convolutional network. The results show that the recognition efficiency and robustness are improved.

Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

- [1] F. Xu, W. Yang, T. Jiang, S. Lin, H. Luo, and G.-S. Xia, "Mental retrieval of remote sensing images via adversarial sketch-image feature learning," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 11, pp. 7801–7814, 2020.
- [2] M. Amani, B. Brisco, S. Mahdavi et al., "Evaluation of the Landsat-based Canadian wetland inventory map using multiple sources: challenges of large-scale wetland classification using remote sensing," *Ieee Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 32–52, 2021.
- [3] H. Ghassemian, "A review of remote sensing image fusion methods," *Information Fusion*, vol. 32, pp. 75–89, 2016.
- [4] K. Li, W. Xie, Q. Du, and Y. Li, "DDLPS: detail-based deep laplacian pansharpening for hyperspectral imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 10, pp. 8011–8025, 2019.
- [5] T. Benzenati, A. Kallel, and Y. Kessentini, "Two stages pansharpening details injection approach based on very deep residual networks," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 6, pp. 4984–4992, 2021.
- [6] R. Sun, *Research and Implementation of Fault Tolerance Enhancement Technology for Deep Neural Network*, University of Electronic Science and Technology of China, Chengdu, China, 2022.
- [7] Z. Ying, G. Li, and W. Gao, "A bio-inspired multi-exposure fusion framework for low-light image enhancement," 2017, <https://arxiv.org/abs/1711.00591>.
- [8] Q. Zhang, Y. Nie, and W. S. Zheng, "Dual illumination estimation for robust exposure correction," *Computer Graphics Forum*, vol. 38, no. 7, pp. 243–252, 2019.
- [9] W. Wang, Z. Chen, X. Yuan, and X. Wu, "Adaptive image enhancement method for correcting low-illuminati on images," *Information Sciences*, vol. 496, 2019.
- [10] Z. Zhou, S. Li, and B. Wang, "Multi-scale weighted gradient-based fusion for multi-focus images," *Information Fusion*, vol. 20, pp. 60–72, 2014.
- [11] X. Fu, D. Zeng, Y. Huang, Y. Liao, X. Ding, and J. Paisley, "A fusion-based enhancing method for weakly illuminated images," *Signal Processing*, vol. 129, pp. 82–96, 2016.
- [12] B. Liang, *Research on the Detection Method of Irregularly Shaped Text Lines Based on Deep Convolution Neural Network*, Xiamen University, Xiamen, China, 2019.
- [13] J. Shi, *Research and Implementation of Driver's License Recognition Technology Based on Neural Network in Complex Background*, Xiamen University, Xiamen, China, 2019.
- [14] H. Liu, *Research on Chinese News Text Classification Based on Deep Learning*, Southwestern University of Finance and Economics, Chengdu, China, 2019.