ISSN: 2414-1895 DOI: 10.6919/ICJE.202205_8(5).0111

Image Recognition of Esophageal Cancer based on ResNet and Transfer Learning

Qiying Ling, Xiaofang Liu

School of Computer Science and Engineering, Yibin 644005, China

Abstract

Based on the image recognition accuracy and other indicators, the feasibility of transfer learning to solve the problem of the small amount of esophageal cancer data sets and incomplete labeling information is discussed. Select 200 esophageal endoscopic images of pathological patients with esophageal cancer and 100 esophageal endoscopic images of unaffected patients. After mixing and scrambled, they are divided according to a certain proportion, and the transferred ResNet network is used for training and learning. The results show that The feasibility of applying transfer learning to esophageal cancer image recognition with a small amount of data is also demonstrated, and the feasibility of deep learning as an auxiliary diagnosis of esophageal cancer images is also demonstrated.

Keywords

Transfer Learning; Esophageal Cancer; Resnet.

1. Introduction

According to the 2020 Global Cancer Report published by the International Agency for Research on Cancer (IARC), a subsidiary of WHO, there will be 19,292,789 new cancer cases worldwide in 2020, of which the incidence of esophageal cancer is as high as 604,100, accounting for the global 3.1% of the total incidence, and cervical cancer tied for seventh. In 2020, there were 9,958,133 new cancer deaths worldwide, and esophageal cancer alone caused 544,076 deaths, accounting for 5.5% of the total, ranking sixth. In addition, the report also showed that Asian countries ranked first in terms of morbidity, mortality, and five-year prevalence rates, accounting for 79.7%, 79.8%, and 78.5%, respectively. China, which is in Asia, accounts for nearly half of the world's new cases and deaths, and many provinces are high-incidence areas of esophageal cancer.

Moreover, various types of traditional cancers rely on doctor's diagnosis, and there will be certain misdiagnosis and missed diagnosis. In recent years, with the rapid development of the computer field, convolutional neural networks have been widely used in image recognition and classification, such as character recognition [1-2], face recognition [3-4], animal and plant recognition [5], medical image pathological classification [6], etc. There are many different models of neural networks in deep learning, which can be divided into LeNet, AlexNet, VGG, GoogLeNet, ResNet and other models according to the development time. Among them, the ResNet model was the champion of the 2015 ImageNet competition, reducing the recognition rate to 3.6%. This result has exceeded the accuracy of normal human eye recognition. The most classic part is the use of the residual module (Residual block) to solve the gradient Diffusion problem. This article will discuss the feasibility of using transfer learning to accurately identify esophageal cancer images. First, it will discuss the principles of transfer learning and the ResNet model, then conduct experimental research, and finally analyze the experimental results and propose revisions.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202205_8(5).0111

2. Theoretical Basis

2.1 Transfer Learning

In order to solve the problem of poor training results due to insufficient training samples in the process of machine learning and deep learning, scholars have proposed the idea of transfer learning [7]. The specific idea of transfer learning is to use sufficient datasets in related fields to train network models, and then transfer the parameters and models to the research field to complete the corresponding tasks [8]. According to whether the samples in the source domain and the target domain are labeled and whether the tasks are the same, the previous transfer learning work can be divided into inductive transfer learning, transductive transfer learning, and unsupervised transfer learning. According to the technology used in transfer learning, transfer learning can be divided into feature selection-based transfer learning algorithm research, feature mapping-based transfer learning algorithm research, and weight-based transfer learning algorithm research. The source domain data and target domain data used in transfer learning tasks are related, although distributed differently. That is, a portion of the training samples in the auxiliary source domain is suitable for learning an effective classification model, and is applicable to the target test samples. In recent years, with the wide application of deep learning, transfer learning has also been applied in various fields. For example, Lin Yu et al. proposed a remote sensing image classification method based on deep transfer transfer learning [9]; Huang Xiaxuan et al. used transfer learning to realize medical images. Data transfer classification [10]; Li Xianguo et al. used transfer learning to realize the detection of conveyor belt breakage of permanent magnet iron remover [11]; Wang Xin et al. used transfer learning to study the classification method of poisonous jellyfish [12] and so on.

The main steps of transfer learning are as follows:

1) Use a large number of marked data sets (source domain) to train the neural network, and perform feature extraction on the source domain images through the convolutional layer and pooling layer at the front of the model;

2) Pre-training model, import the trained model into the target task, and reconstruct the classification layer by customizing the fully connected layer;

3) Fine-tune, freeze the network parameters of the previous layers, train with the target domain image, record the parameters of the forward propagation through forward propagation, apply the trained model to the target task, and complete the transfer learning;

Parameter fine-tuning can solve the problem that the feature parameters of the pre-trained neural network model do not match the task parameters in the target domain, and is the most important step in transfer learning.

2.2 Convolutional Neural Network

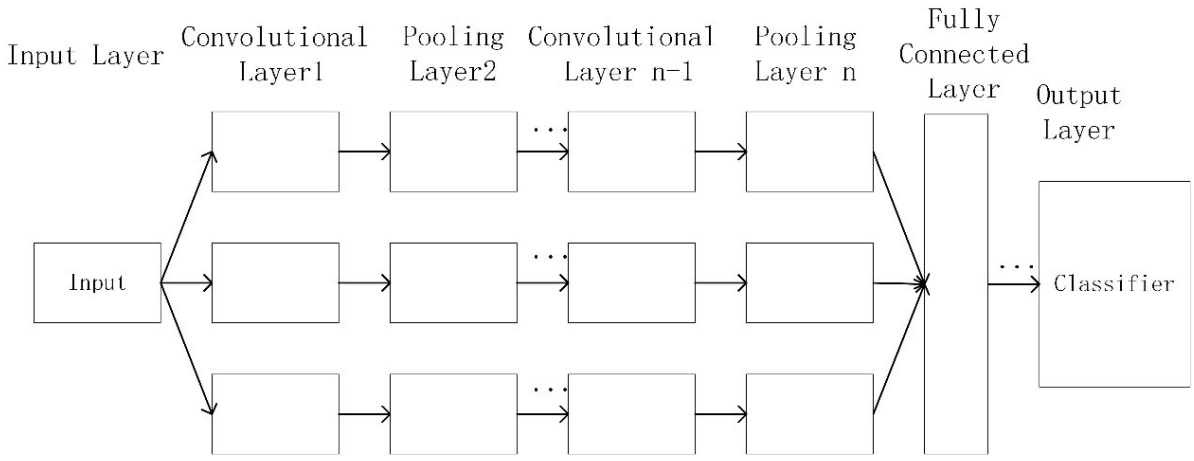

Figure 1. Convolutional Neural Network Architecture

ISSN: 2414-1895 DOI: 10.6919/ICJE.202205_8(5).0111

Convolutional neural networks consist of one or more convolutional layers, pooling layers and fully connected layers. Compared with other neural network structures, convolutional neural networks can achieve better results in terms of images, etc. Post-Layer Deep Neural Networks, Convolutional Neural Networks have fewer parameters to consider, making them the most popular deep learning architecture. The basic network structure of convolutional neural network is shown in Figure 1.

The convolutional layer contains several feature surfaces, each feature surface contains multiple neurons, each neuron is connected to the previous feature surface through a convolution kernel, and the convolution kernel can be regarded as a weight matrix, and the . The weights of the same eigenface are shared. The output result is calculated from the weighted and biased values of the local area of the input feature surface.

The pooling layer is arranged behind the convolutional layer. The convolutional layer is the input of the pooling layer, and the neurons of the feature surface of the pooling layer are also connected to the local area of the input feature surface, which is obtained by reducing the dimension of the feature surface. Scale-invariant properties of features. Pooling methods generally include max pooling and average pooling.

The input features are passed alternately through multiple convolutional layers and pooling layers. The convolutional neural network classifies the extracted features through the fully connected network, and finally obtains the probability distribution of the current sample through the Softmax layer.

2.3 ResNet Model

When it comes to deep learning, the easiest thing to notice is the depth. It has always been generally believed that the deeper the depth, the better the recognition and classification effect of the model. However, studies have shown that when the depth reaches a certain level, the performance of the model will stop increasing, or even drop significantly. This is because with the deepening of the model depth, the training difficulty of the neural network will also increase. Between the deep neural networks, when the gradient information is passed forward, the gradient disappears close to 0, and the number of network layers The deeper it is, the greater the probability of gradient disappearance. The residual learning proposed by He Kaiming [13] et al. adds a skip connection between some inputs and outputs in the deep network to form a residual module, so that the input information can be directly passed to the next layer [14], so as to realize the layer number fallback mechanism.

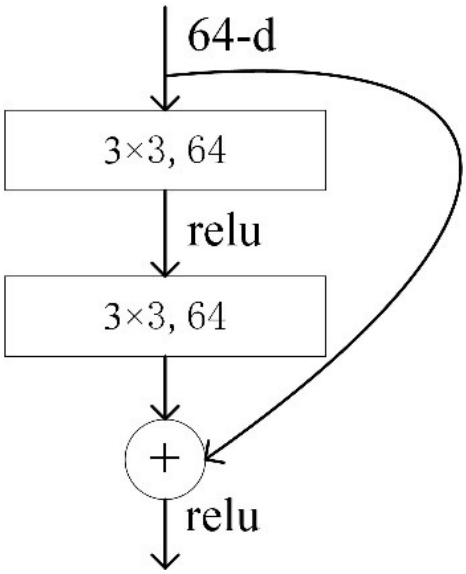

Figure 2. Shallow Residual Module

ISSN: 2414-1895 DOI: 10.6919/ICJE.202205_8(5).0111

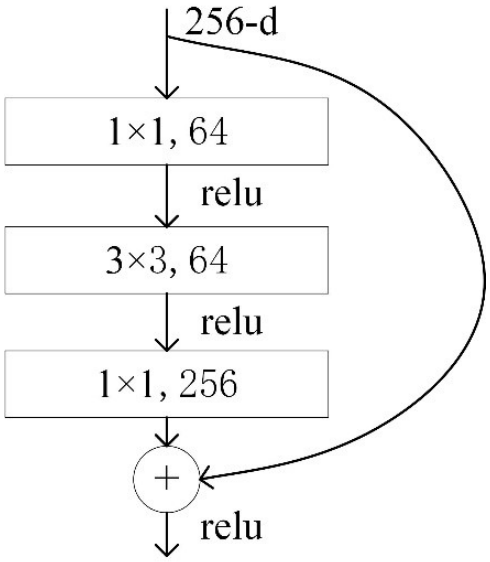

Figure 3. Deep Residual Module

ResNet uses two residual modules, one corresponding to the shallow layer, as shown in Figure 2, and the other corresponding to the deep layer, as shown in Figure 3. The purpose is to reduce the amount of computation and parameters. Table 1 shows five commonly used ResNet models with depths of 18, 34, 50, 101, and 152. It is not difficult to find that all deep ResNet models are composed of conv1, conv2_x, conv3_x, conv4_x, and conv5_xz. Each of the convolutional modules has certain differences.

Table 1. ResNet different types of structures

layer name	output size	18-layer	34-layer	50-layer	101-layer	152-layer
conv1	112×112	7×7, 64, stride 2				
conv2_x	56×56	3×3 max pool, stride 2				
		$\begin{bmatrix} 3\times3, 64 \\ 3\times3, 64 \end{bmatrix}\times2$	$\begin{bmatrix} 3\times3, 64 \\ 3\times3, 64 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 256 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 256 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 256 \end{bmatrix}\times3$
conv3_x	28×28	$\begin{bmatrix} 3\times3, 128 \\ 3\times3, 128 \end{bmatrix}\times2$	$\begin{bmatrix} 3\times3, 128 \\ 3\times3, 128 \end{bmatrix}\times4$	$\begin{bmatrix} 1\times1, 128 \\ 3\times3, 128 \\ 1\times1, 512 \end{bmatrix}\times4$	$\begin{bmatrix} 1\times1, 128 \\ 3\times3, 128 \\ 1\times1, 512 \end{bmatrix}\times4$	$\begin{bmatrix} 1\times1, 128 \\ 3\times3, 128 \\ 1\times1, 512 \end{bmatrix}\times8$
conv4_x	14×14	$\begin{bmatrix} 3\times3, 256 \\ 3\times3, 256 \end{bmatrix}\times2$	$\begin{bmatrix} 3\times3, 256 \\ 3\times3, 256 \end{bmatrix}\times6$	$\begin{bmatrix} 1\times1, 256 \\ 3\times3, 256 \\ 1\times1, 1024 \end{bmatrix}\times6$	$\begin{bmatrix} 1\times1, 256 \\ 3\times3, 256 \\ 1\times1, 1024 \end{bmatrix}\times23$	$\begin{bmatrix} 1\times1, 256 \\ 3\times3, 256 \\ 1\times1, 1024 \end{bmatrix}\times36$
conv5_x	7×7	$\begin{bmatrix} 3\times3, 512 \\ 3\times3, 512 \end{bmatrix}\times2$	$\begin{bmatrix} 3\times3, 512 \\ 3\times3, 512 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 512 \\ 3\times3, 512 \\ 1\times1, 2048 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 512 \\ 3\times3, 512 \\ 1\times1, 2048 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 512 \\ 3\times3, 512 \\ 1\times1, 2048 \end{bmatrix}\times3$
	1×1	average pool, 1000-d fc, softmax				
FLOPs		1.8×10^9	3.6×10^9	3.8×10^9	7.6×10^9	11.3×10^9

3. Experimental Study

3.1 Datasets and Preprocessing

The data set adopts the esophageal endoscopic image data set published on Kaggle, and selects 100 esophageal endoscopic images of non-pathological patients and 200 esophageal endoscopic images of pathological patients in the data set. The endoscopic image of the esophagus is shown in Figure 4.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202205_8(5).0111

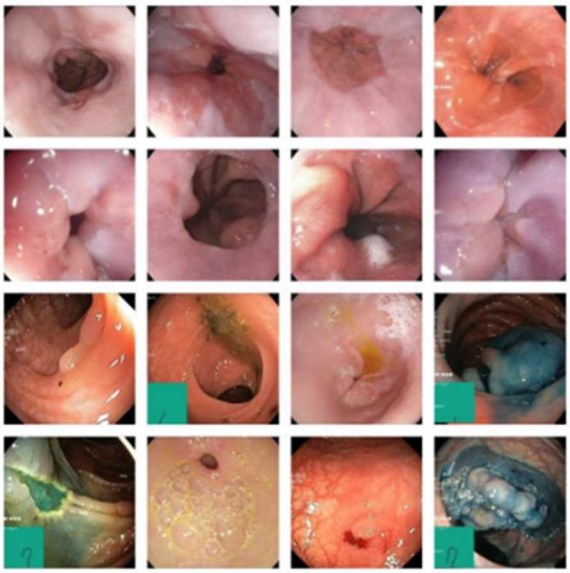

Figure 4. Esophageal Endoscopy Image

As shown in Table 2, the data set is divided into training set (training), validation set (validation), and test set (testing) according to the ratio of 6:2:2, and the esophageal endoscopy images of pathological patients are used as positive samples. The label is 1, and the esophagoscopic images of non-pathological patients are used as negative samples, and the label is 0. Considering the difference in light source brightness, operating equipment, operating habits, and disease cases during the acquisition of esophageal endoscopic images, the images were rearranged, numbered, cropped, data enhanced, and transformed. The research shows that the features learned by the neural network are invariant when the image is panned and zoomed. OpenCV is used to read the image from the disk, and each image is renumbered, denoised, and cropped to a size of 224*224. The pixel values are batch normalized to be between[-1,1].

Tabel 2. Distribution of Experimental Data

Category	Training Set	Validation Set	Test Set
Esophageal Cancer	120	40	40
Non-esophageal Cancer	60	20	20

3.2 Load the Pretrained Network

This experiment mainly uses PaddlePaddle2.1.2+python3.8 version as the basic framework of this study, in which the ResNet network model can be loaded directly. After loading the network model, use the ImageNet dataset for pre-training, and freeze and fine-tune the trained ResNet model. When the last 3 layers of the pretrained network net are configured for 1000 classes.

Then these 3 layers must be fine-tuned for the new classification problem. First extract all layers except the last three layers from the pre-trained network, then add the fine-tuned fully connected layer, Softmax layer and classification output layer to migrate to the new classification task, and finally specify a new fully connected according to the new data options for the layer, set the fully connected layer to be the same size as the number of classes in the new data.

3.3 Experimental Results and Analysis

The fine-tuned network classifies the validation images and obtains an accuracy of 92.32%. Compared with the accuracy rate of 80.12% obtained by direct training with the same dataset, there

ISSN: 2414-1895 DOI: 10.6919/ICJE.202205_8(5).0111

is a significant improvement, which proves the feasibility of transfer learning in the field of medical images where small datasets are the basic norm.

The results of this experiment may be relatively good because of the simple classification. Considering that the actual disease situation may be more complex, ResNet can be improved by adding training data, refining data classification, and further optimizing model parameters and training parameters. Classification accuracy and generalization ability of the model on esophageal endoscopic images of esophageal cancer.

4. Conclusion

Considering that the current diagnosis of esophageal cancer mainly relies on doctors, it is slightly more stressful for doctors to announce work, and there is a possibility of misjudgment or missed judgment. In the field of image classification, the convolutional neural network has a relatively simple implementation process and is easy to operate. It can be used as an auxiliary detection method for early telemedicine diagnosis, which can not only improve the efficiency of diagnosis, but also reduce the pressure on doctors.

This paper mainly analyzes the performance and feasibility of the application of transfer learning in small datasets of esophageal cancer images. The experimental results show that it is feasible to use transfer learning to identify esophageal cancer images, and the accuracy rate can reach 92.32% in the test set. The experiments also demonstrate that CNN can be used as an auxiliary diagnostic tool for esophageal cancer. In future medical image research, small datasets in the field of medical images are the basic norm, and there are generally certain categories and imbalances between positive and negative samples. Transfer learning can be considered to obtain higher accuracy and generalization ability.

Acknowledgments

This work is supported by Funded by Sichuan Science and Technology Program, Project number: 2017GZ0303. Sichuan Academician (Expert) Workstation Fund Project, Project number: 2016YSGZZ01. Special Funding for High-level New Talent Training, Project number:B12402005. Talent Introduction Project of Sichuan University of Light Chemical Industry, Project number: 2021RC16. Supported by The Innovation Fund of Postgraduate, Sichuan University of Science & Engineering.

References

[1] Wang Dong. Application Research of Artificial Intelligence OCR Technology[J].Electronic Technology and Software Engineering, 2022(01):122-125.

[2] Zhang An. Research on Text Recognition Based on Tesseract [D]. Nanjing University of Posts and Telecommunications, 2021. DOI: 10.27251/d.cnki. gnjdc.2021.000439.

[3] Shi Jie. English online test system based on face recognition technology [J]. Information Technology, 2022(02):20-24.DOI:10.13274/j.cnki.hdzj.2022.02.004.

[4] Jiang Tianshui, Wang Jianguo.Ground Penetrating Radar Road Recognition Technology Based on 3D Face Recognition of HD Camera[J].Modern Radar,2022,44(02):64-68.DOI:10.16592/j.cnki.1004-7859. 2022.02.010.

[5] Zhang Xueqin, Chen Jiahao, Zhuge Jingjing, et al. Fast plant image recognition based on deep learning [J]. Journal of East China University of Science and Technology (Natural Science Edition), 2018,44(6): 887-895.

[6] Zhang Zezhong, Gao Jingyang, Lv Gang, et al. Classification of gastric cancer pathological images based on deep learning [J]. Computer Science, 2018, 45(2): 263-268.

[7] Pan SJ,Yang Q.A survey on transfer learning[J].IEEE Trans Knowl Data Eng,2010,22(10):1345-1359.

[8] Zhuang Fuzhen, Luo Ping, He Qing, et al. Research progress of transfer learning [J]. Journal of Software, 2015, 26(1): 26-39.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202205_8(5).0111

[9] Lin Yu, Zhao Quanhua, Li Yu. A Remote Sensing Image Classification Method Based on Depth Transfer Transfer Learning[J].Journal of Earth Information Science,2022,24(03):495-507.

[10] Huang Xiaxuan, Huang Tao, Yuan Shiqi, He Ningxia, Wu Wentao, Lv Jun. Implementation of transfer learning of medical imaging data based on MATLAB [J]. Medical News, 2022,32(01):33-39.

[11] Li Xianguo,Liu Xiao,Feng Xinxin.Break detection of conveyor belt of permanent magnet iron remover based on transfer learning[J].Journal of Tianjin University of Technology,2022,41(01):66-72.

[12] Wang Xin, Wen Zehua, Ren Jiale, Han Yiyuan.Research on the classification method of poisonous jellyfish based on transfer learning[J].Journal of Lanzhou Vocational and Technical College,2022, 38(01): 67-71.

[13] He K,Zhang X,Ren S,et al.Deep residual learning for image recognition[A].Proceedings of the IEEE conference on computer vision and pattern recognition[C].Las Vega (US):IEEE,2016,770-778.

[14] CAO F K, BAI T, XU X L. Vehicle detection and classification based on highway monitoring video[J]. Computer Systems & Ap- plications, 2020, 29(10) : 267-273.