Research on Lung Medical Image based on Convolution Neural Network Algorithm

Yipeng Xu1,*, Yiming Zhu2, Zhongrun Chen3

1School of Mathematics and Science, TianGong University, Tianjin, 300387
2Beijing Institute of Technology, Beijing, 100000
3School of Materials Science and Engineering, Xiangtan University, Hunan, 411105

*Corresponding author: yipengxu@tiangong.edu.cn

Abstract. In the field of medical image processing, the convolution neural network (CNN) algorithm in depth learning solves many problems in machine vision. Aiming at the problems of low resolution of medical images and easy misjudgment of artificial recognition, in order to help medical workers to better identify the lesion area, this paper discusses the structure optimization, training and prediction ability of CNN, and trains the convolution neural network to judge the lesion area in the image and get more accurate discriminant results on lung images, further verify the high efficiency, accuracy and reliability of convolutional neural network algorithm in medical image field.

Keywords: Convolutional neural network; Medical image processing; Supervised learning; Fine-grained classification.

1. Introduction

Machine learning technology is becoming more and more mature in the field of machine vision, and deep learning algorithm [1] is gradually applied to industrial production and service industry, and it has performed well in some artificial intelligence tasks. It has proved to be good at discovering structures in intricate high-dimensional data, and therefore applies to many fields of science, business and government. Medical aspects, showing higher efficiency than other machine learning algorithms in predicting the activity of potential drug molecules and predicting the effects of gene mutations on gene expression and disease.

In the medical field, deep learning [2] can be used for the classification of medical images, the comprehensive judgment of various indicators of patients and the diagnosis of etiology.

Medical image is a special kind of image, which is highly dependent on environment, and has many kinds of images and is difficult to observe small defects. Each biological individual has variability, so it is difficult to give accurate and objective criteria for special individuals, deep learning algorithm can effectively solve this problem. Deep learning is good at finding statistical laws from a large number of statistical information, which can learn real characteristics from different individual cases, so as to make accurate judgment or give location, classification information.
Because of its strong statistical inference ability and low demand for manual marking, deep learning is bound to achieve many successes in medical images in the future. Currently, new learning algorithms and architectures being developed for deep neural networks will accelerate this process.

2. Overview of the Deep Network

Convolutional Neural Network (Convolutional neural networks, CNN) is a class of data depth network structures used to process data in the form of multiple arrays, such as color images consisting of three two-dimensional arrays (which contain pixel intensities in three color channels). During processing, many data forms appear in the form of multiple arrays. Behind the CNN are four key ideas: local connectivity, shared weights, pooling, and using multiple layers.

The typical CNN architecture is composed of a series of layers, as shown in figure 1, the classical LeNet-5 network structure. The previous layers consist of two types of layers: convolutional layer and pooling layer. There are multiple convolution kernels in the convolution layer. Each convolution kernel is operated by sliding convolution in the feature map by connecting the feature map of the previous layer to generate a new feature map. This can be seen as the mapping operation of the upper-level feature to the lower-level feature. These locally weighted results also have to be activated by the nonlinear activation layer to achieve the purpose of increasing the nonlinearity, so as to better simulate the inference process of the natural system. There are two reasons for this architecture: first, in array data such as images, local value groups are usually highly correlated to form unique local patterns that are easy to detect; secondly, the local statistics and positions of images and other signals are invariant.

The function of pooling layer is to merge semantically similar features into one feature. Since the relative positions of the features that make up a topic may vary, the subject can be reliably detected by coarse grinding the positions of each feature. A typical pool cell computes the maximum of the local patches of the cell in a feature graph. Adjacent pooling units obtain inputs from patches that move multiple rows or columns, thereby reducing the dimension of the representation and obtaining invariance to small moves and distortions. The deep neural network takes advantage of many natural signals and is a characteristic of the composite hierarchy. For example, in the image, the local combination of the edges forms a pattern, the pattern is combined into a part, and the part forms an object. Pooling operations minimize changes in the position and appearance of elements in the current layer.

3. Fine-grained recognition algorithm

3.1. R-CNN

R-CNN can not only improve the quality of candidate bounding boxes, but also extract advanced features. Its work can be divided into three stages: generating candidate regions, CNN-based feature extraction and classification and localization.

R-CNN employ selective search to generate 2k region proposals for each picture, selective search methods rely on simple bottom-up grouping and saliency cues, which quickly provide more accurate
candidate boxes of any size and effectively reduce search space. For classification, R-CNN uses the area proposal to score, then adjust the score area by bounding box regression, and filter with non-maximum suppression to produce the final bounding box for retaining the target position.

While R-CNN is more accurate than traditional methods and it is important to apply CNN to real object detection, there are still some shortcomings.

1. The existence of a fully connected layer fixed the size of the input image, which directly results in the entire CNN of each assessment area requiring recalculation and a large amount of test time;
2. R-CNN training is done in multiple stages. First, the proposed convolutional network is fine-tuned. The softmax classifier learned by fine-tuning is then replaced by SVM. Finally, the bounding box regressors are trained;
3. The training period is long and the space is large. Features proposed to be extracted from different regions and stored on disk take up a lot of memory.
4. Although selective search can generate region proposals with high recall rates, the regions obtained still have redundant parts and the process takes a long time. Many methods have been proposed to solve these problems. DeepBoxl and SharpMask try to reorder or refine regional proposals before extracting features to remove redundant information and obtain a few valuable suggestions. In addition, there are some improvements that can solve the problem of inaccurate positioning. Saurabh Gupta et al. improve the target detection of RGB-D images with semantic rich image and depth features. The combination of target detector and superpixel classification framework has achieved gratifying results in semantic scene segmentation task.

3.2. SPP-net

To solve the problem of target loss or distortion caused by the change of target proportion, He et al. considered the theory of spatial pyramid matching (SPM) and proposed a novel CNN architecture called SPP-net. SPM need finer scales to divide images into multiple parts and aggregate quantized local features into intermediate representations.

Unlike R-CNN, feature maps of the fifth convolutional layer (Conv5) are SPP-net reused in target detection to project regions of any size proposed to feature vectors of fixed length. The feasibility of the reusability of these feature maps involves not only the enhancement of the local response but also its spatial location. SPP-net can not only obtain better results by estimating the proposed size of different regions, but also improve efficiency during testing by sharing SPP layers between different regions and calculating costs in advance.

3.3. Fast-R-CNN

Compared with R-CNN, although SPP-net networks have improved a lot in accuracy and efficiency, there are still some problems. SPP-net uses almost the same multi-level operation as R-CNN, requiring additional storage space. Besides, the convolutional layer before the SPP layer cannot update parameters using the fine-tuning algorithm, which will lead to a decrease in the accuracy of the deep network.

To address the above problems, Girshick introduces multiple task losses and proposes a novel convolutional neural network called Fast R-CNN. Fast-R-CNN at the last layer of the convolution, region of interest (Region of Interests, ROI) pooling is used to generate feature maps with fixed dimensions. In addition, the boundary regression is added to the CNN network for training using a multi-task loss function.

Fast R-CNN whole image of the system is processed by convolution layer to generate feature map, then a fixed length feature vector is extracted from the region proposal of each region of interest. Area of interest layer is special SPP layer, it has only one pyramid. It feeds each feature vector into a series of fully connected layers and finally into two equivalent output layers. One output layer generates the softmax probability of all C+ class 1 (C target class plus a "background" class), and the other layer encodes the position of the encoded bounding box with four numerical values. All parameters in these
processes (except those proposed in the generated region) are optimized by the end-to-end multi-service loss function.

3.4. One-stage-based models

The two-stage-based framework consists of several stages, region proposal generation, feature extraction using CNN, classification, and bounding box regression, which are usually trained separately. Even in recent end-to-end network Faster R-CNN, alternative training is still required to obtain shared convolution parameters between the region proposed network and the detection network. Therefore, the time taken to process different parts becomes the bottleneck of real-time target detection. A one-step framework based on global regression/classification can be directly mapped from image pixels to bounding box coordinates and classification probabilities, reducing time overhead. This paper presents several important frameworks, namely, YOLO and SSD.

4. Lung Image Recognition Based on Convolutional Neural Network

4.1. Training of convolutional neural networks

Using supervised learning to train neural networks to have the ability to judge whether it is shadow or not. The main steps are as follows:

1) Design neural networks. Due to the low resolution of the image itself, the judging network input can be set to 32*32 size. Because the function is single, it means that the network faces less image change and the scene is single, so the parameter quantity of convolution and other operations is not too much. Imitate the LeNet network and simplify the parameters.

2) The size of the image cut by the normal image and the shaded image is prepared depending on the size of the shadow to be judged. After cutting the image, the small graph is divided into two categories, and the neural network is trained by image and label. If there is a need to judge other lesion images, modify the number of channels in the last layer of the neural network and prepare the corresponding multi-class data small map.

3) After a round of training, the prediction value is used to predict a batch of untrained images. At this time monitor the correct rate of network judgment and loss function size. If the network accuracy rate has converged to a high level, then stop training, save parameters.

4.2. Algorithm flow

After getting the lung picture, the target is to detect the shadow area. The algorithm flow is as follows:

1) Load trained networks;

2) The effective region of the image is initially segmented and the network judges it after segmentation;

3) If a region is found to have a normal probability below the threshold, the region is further segmented and then judged to be fine-grained;

4) To collect the results of the judgement after dividing it to the minimum scale;

5) Repeat (2)~(4) steps until all areas of the image are covered and finally indicate the detected area.

The classification accuracy of small plots in the test set network reached 95.05%. The last layer of the network is a fully connected layer with a number of channels of 2, resulting in a classification result that is or is not a target. After the operation of the Softmax layer, two numbers with a sum of 1 can be obtained, which can be seen as the judgment probability. In practical application, we can also set a threshold for these two probability values to control the severity of judging shadows. As shown in figure 2, when the correct rate of the character judgment shadow is set to 0.68, the accuracy is better. If adjusted below 0.42 or above 0.87 will cause false detection.
5. Concluding remarks
In this paper, the construction, learning and training of convolutional neural networks in deep learning are described in detail, and the processing methods of similar data and the algorithm flow to improve the fine-grained classification ability are given in the background of lung angiography. The system direction which can be used in medical image recognition is also given, so that the advantages of deep learning can be applied in medical image recognition.

Deep learning demonstrates better processing power than other machine learning algorithms and the potential is still being explored. In the future, there will be more standard neural network training mode, deeper network level and better training methods, which can be applied to more machine vision field.

References
[1] Chunhua Wang, Dong Han. Adaptive Control Depth Learning and Knowledge Mining Image Classification [D]. Journal of Shenyang University of Technology, 2018, 40(3): 334-339.
[2] Jiang Guiping, Qin Wenjian, Zhou Shoujun, et al. Current Situation of Medical Image Segmentation and Development [D]. Journal of Computer Science, 2015, 38(6):1222-1242.
[3] An Qiangqiang, Zheng Min. Image Recognition Based on Deep Learning [D]. Automation and Instrumentation, 2018, (3): 115-118.
[4] Huang Liqun, Ding Xuesong, Zhang Buzhong, et al. A prediction model of sequence specificity DNA deep learning [J]. Microcomputer Systems, 2018, (11): 2424-2427.
[5] Chen Shihui, Liu Weixiang, Qin Mani, et al. Advances in Computer-Aided Diagnosis of Cancer Based on Deep Learning and Medical Images [J]. Journal of Biomedical Engineering, 2017, (2): 314-319.
[6] Zhang Qiaoli, Zhao Di, Chi Xuebin. a review of medical imaging diagnostics based on deep learning]. Computer Science, 2017, 44(s2): 1-7.
[7] Pan Hao, Wang Zhao. deep learning based lung cancer cell detection method study]. Automation and Instrumentation, 2017, (3): 196-197,200.
[8] Drakakis K,Rickard S,Frein R D,et al.Analysis of financial data using nonnegative matrix factorization[J]. International Mathematical Forum, 2008, 3 (38) : 1853-1870.
[9] Ju B,Qian Y T,Ye M C.Col labor ative filtering algorithm based on structured projective nonnegative matrix factorization [J]. Journal of Zhejiang University: Engineering Science, 2015, 49(7): 1319-1325.