ATLANTIS PRESS

Advances in Computer Science Research, volume 86

International Symposium on Communication Engineering & Computer Science (CECS 2018)

Computer Vision Applied in Medical Technology: The Comparison of Image Classification and Object Detection on Medical Images

Haotian Yan

International School, Beijing University of Posts and Telecommunications, Beijing 100876, China.

yanhaotian@bupt.edu.cn

Abstract. Image classification and object detection are two computer vision techniques that are currently commonly used. In this paper, convolutional neural network (CNN) and region-based CNN (RCNN) are used as examples to analyze and compare image classification and object detection. This paper will analyze the architectural characteristics and application scenarios of these two algorithms and analyzes the different characteristics of these two technologies in medical technology applications. CNN is an infrastructure classification algorithm, and image classification tasks are more common in medical image processing. RCNN is the development of CNN. Object detection technology can directly detect the presence and location of the lesion in medical images with RCNN. Combining the algorithms of the two techniques can also achieve some more complex image processing goals.

Keywords: Deep learning, Convolutional Neuron Network, R-CNN, Medical Technology, Architecture, Comparison.

1. Introduction

Computer vision is one of the fastest growing research areas in the field of artificial intelligence. Deep learning is one of the most dependent techniques of computer vision, and it is a common technique based on artificial neural network, which is used to process the features of images with multi-level network architecture. The convolutional neural network applied in image classification is one of the most representative deep learning algorithms, and the most classic object detection algorithm Region-CNN(RCNN) series (RCNN, Fast RCNN, Faster RCNN) is also based on deep learning technology to achieve classification and localization of objects in the image[1,2,3,4].

Computer vision has penetrated into every corner of modern intelligent life. Computer vision has been extensively employed in the fields of autonomous cars, education, and military. In fact, scientists have been working hard to expand the application of computer vision in the field of medical technology. In 2017, Andrew Ng's team proposed a new network architecture called ChexNet. ChexNet can automatically diagnose pneumonia through chest X-rays, and the diagnostic accuracy exceeds that of radiologists [5].

Image classification and object detection are two primary mission objectives in the field of computer vision. On the basis of these two methods, this paper will discuss the application of these two methods' algorithms in the medical field.

In terms of content, this paper will mainly compare the algorithms' feature and basic framework of image classification and object detection to analyze the application characteristics of these two methods. This paper will mainly use the CNN and RCNN algorithms as examples to make macro and more specific comparisons. In terms of structure, this paper will mainly compare and analyze these two classification methods from the historical development, macro structure, specific functions, characteristics, and application of these two algorithms.

2. Literature Review

The typical representation of CNN as a deep artificial neural network has largely influenced other artificial neural network frameworks for deep learning. In this paper, CNN and R-CNN are introduced as typical algorithms for image classification and object detection in the following sections.

2.1 Image Classification

Image classification is a technique that utilizes a CNN to classify images according to image semantics. Yann Lecun et.al first proposed a convolutional neural network framework in 1998, but until the 2012 ImageNet Challenge convolutional neural networks made a huge breakthrough in image classification [1,6]. Subsequent improvements to CNN, such as VGGNet, ZFNet, and ResNet, performed magnificently on image classification tasks[7,8,9].

2.2 Object Detection.

Object detection is a method of recognizing the class of a specific object in an image by combining image classification and object localization. In 2014, Ross Girshick et.al proposed the RCNN network framework, which combines CNN with regional features to achieve the function of object detection [2]. In order to make the precision higher, the training and testing speed faster, and the improved object detection algorithms such as SPPnet, Fast RCNN, Faster RCNN are created, which are all enhanced algorithms of RCNN[3,4,6,10]. This paper will introduce the architecture and features of RCNN and its improved algorithms in the third part.

Object detection and image classification are both dominant technology of computer vision. Driverless cars require Faster RCNN to perform real-time object detection of obstacles appearing in the front area. Botanists train the CNN network to classify leaf images in order to deduce the plant growth environment. In the field of medical technology, CNN can replace the manual classification of medical images, such as classification of skin photos of skin patients. RCNN can identify specific objects in medical images, such as pinpointing tumors, colon polyps, hydronephrosis, etc., and finding symptoms in the image to diagnose the patient's disease. In this paper, we take the algorithms CNN and R-CNN as examples to compare the algorithm structure and function of the image classification and object detection on medical images.

3. Comparison

The CNN and RCNN algorithms are two typical algorithms for image classification and object detection. This section will take CNN and RCNN algorithms as examples to compare and analyze the structure and function of the algorithms.

3.1 The Architecture of CNN

CNN and some of the deep networks built on it are the most commonly used algorithms in today's image classification tasks. The architecture of these algorithms is mainly composed of CNN and classifier. CNN is used to extract feature vectors from different positions of the image. These features are then processed by the classifier to classify the images. The most common classification method is taking use of the softmax function, softmax computes the probability of that each category is compatible with image. GoogLeNet, VGGNet, AlexNet are all constructed based on a similar structure [6,7,8].

Taking AlexNet as an example, the main components of AlexNet are convolutional computing layers, pooling layers and fully connected layers, as shown in Figure 1 [6]. The convolutional computing layer is the core of the whole convolutional neural network. Filters, as known as convolution kernel, are superimposed. Each filter is a small matrix that performs inner product operations on the data of a particular window in the image, a so-called convolution calculation. The maximum pooling layer separates the data (matrix) obtained from the convolution calculation into several small bins of equal size, and then selects the maximum value of each small block to form a new matrix, eliminating redundant data. At the end of the CNN, the fully connected layer is usually set, and the feature vector processed by the fully connected layer is converted by the softmax function finally into a probability distribution which assign a probability to each category.

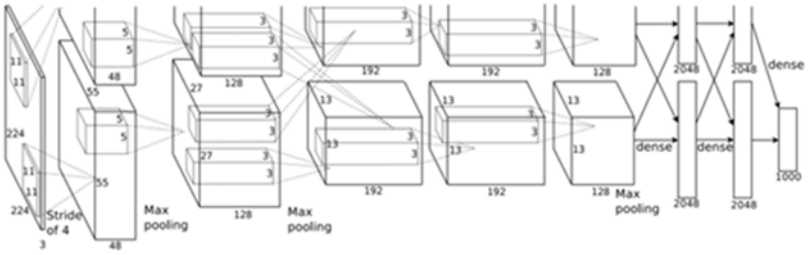

Fig 1.This is a schematic diagram of AlexNet's structure. In the figure we can see the convolutional layer, max-pooling, fully connected layer [6].

3.2 The Architecture of RCNN

RCNN is an algorithm that combines CNN and region features to achieve object detection goals. The structure of R-CNN is shown in Figure 2 [2]. There are two differences between RCNN and CNN architecture: 1. RCNN first extracts 2000 region proposals by using selective search, and then CNN calculates features for each region proposal. 2. Use the linear support vector machine (SVM) to classify the region proposal, and then use the bounding-box regressor to confirm the location of the object.

Extracting the region proposal is realized by sliding the window of the selective research algorithm. Because CNN requires images of the same size as input, the region proposal is cropped or wrapped to the same size image before being entered into the CNN. CNN calculates features for these same-size images, and classifier then categorizes each region proposal. As for the method of classification, the RCNN train a binary SVM for each category. Because the object detection needs to obtain the objects' location, the RCNN uses the bounding-box regressor to eliminate the redundant region proposal and find the most accurate location of the object.

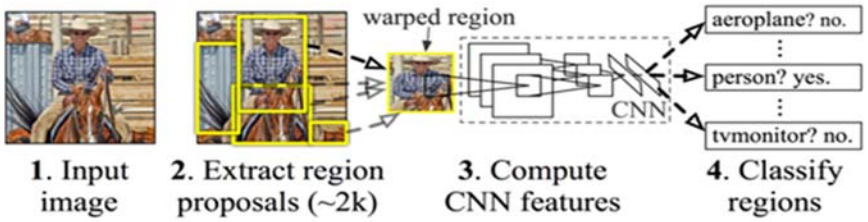

Fig 2. In the second step, the RCNN extracts 2000 region proposals for the entire image, and the CNN of the third step is used to calculate features for these region proposals [2].

3.3 The Development of RCNN

Established on CNN, RCNN add the phase of extracting region proposal and achieve the objectiveness of implement object detection with CNN. However, since the feature is calculated for each region proposal, the calculation of the CNN is extensive, resulting in a slow speed while testing and training. In addition, cropping and wrapping also have a negative impact on accuracy. Fast RCNN and Faster RCNN have improved the architecture of CNN, and its corresponding structure is shown in Figure 3 and Figure 4 [3,4]. Fast RCNN replaces the last max-pooling of CNN with the RoI pooling layer, and replaces the last layer of FC and softmax with two parallel structures like Figure 3 (connecting the softmax and bounding box regressor with a fully connected layer). From Figure 3, it can be seen that the Fast RCNN is a single-stage, which is more efficient than the RCNN that is a multi-stage processing. Because the approach to extract the region proposals adopted by Fast RCNN is still selective research, which limits the increase of testing and training speed, Faster RCNN constructs a new network region proposal network (RPN) instead of the selective search algorithm to extract the region proposal. The RPN is based on CNN. Figure 4 illustrates the architecture diagram of Faster RCNN. It can be found that the convolution part of RPN and CNN are shared, ensuring the two networks can share the calculation results and improve the speed and precision. Faster RCNN is

nowadays the most commonly used object detection algorithm as well as the fastest and most accurate algorithms in the RCNN series algorithms.

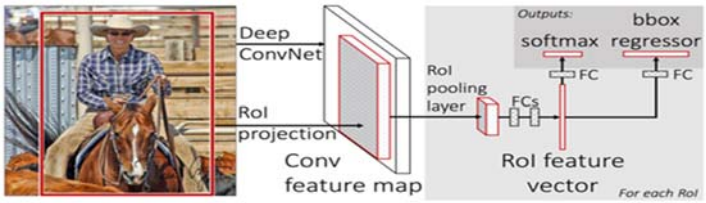

Fig 3. The input of Fast RCNN is the image and its regions of interests (RoI).

The RoI pooling layer is unified into the same size for these RoI feature vectors, input into two parallel branches, where softmax gets the probability distribution column of each category. The bbox regressor generates a four-dimensional vector that locates the position of the RoI [3].

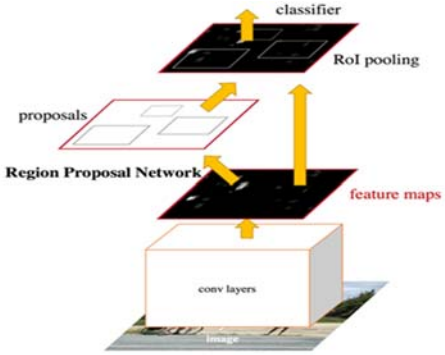

Fig 4. The RPN slides over the feature map. The RPN also contains a classification layer and a regression layer to find the appropriate region proposal [4].

4. Discussion of Application and Function

This part will compare and analyze image classification and object detection from the perspective of processing medical image, and will also introduce the application features and functions of the two techniques at the same time.

The difference between the image classification and the object detection at application level can be summarized that the classification task is relatively more commonplace than object detection, because not only the image will involve the classification task, the text and the voice can be regarded as objects for classifying and these classification tasks usually adopt specific algorithms created on the CNN. In fact, the CNN algorithm for image classification is an algorithm that is more analogous to an infrastructure and fundamental framework. It is applied to classify images at the image level. The RCNN algorithm for object detection tasks is more concrete and complex than CNN, which process at a level of part of the image (region proposal). In the field of medical technology, these two algorithms also conform their particular characteristics at the application level. Two examples will be given below to illustrate the distinctions.

Faster RCNN is the most advanced and practical algorithm in RCNN series for object detection. In 2017, Jane Hung and Anne Carpenter used Faster RCNN to achieve object detection of malaria cells, which can be found in microscopic photographs of blood smears [11]. The researchers selected microscopic blood smears from three of the four patients as training datasets and microscopic blood smears from the other patient for testing. The architecture of this algorithm, shown in Figure 5, is a two-stage approach consisting of a Faster RCNN and an AlexNet. The Faster RCNN performs object detection on the image, and after objects being detected, they are marked as red blood cells (RBC) or other cells (because the red blood cells occupy an extremely high percentage in the blood). Then the labeled images are input into AlexNet, which classifies the non-red blood cells into cell infected with malaria and non-infected cells. Previous malaria detection methods often use a classification method of machine learning, which lacks universality and is complicated to replicate. Through the experiment,

researchers found the accuracy of this traditional algorithm is 50%, while the accuracy of the two-stage algorithm using Faster RCNN is 98%, and the accuracy of human experts is 72%.

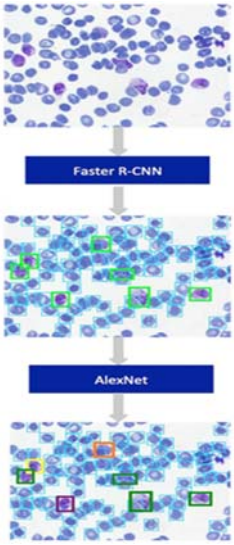

Fig 5. The images processed by Faster RCNN were labeled with non-RBC, and AlexNe was used to classify labeled cells [11].

Jia Ding et.al built a two-stage computer aided system (CAD) to detect lung nodules using the improved Faster RCNN and 3D CNN, and won the first place in the 2016 Lung Nodule Analysis Challenge [12]. The first stage of the CAD is an upgraded Faster RCNN with the axial slices of the Computer tomography (CT) image as input. As shown in Fig 6, in the last layer of the CNN part of the Faster RCNN, a deconvolutional layer is added for the difficulty in extracting the RoI of nodules caused by the inexplicit extraction feature; and when the RPN network slides on the feature map, the number of anchors per window is raised up to 6 in order to accommodate different probable sizes of nodules. The improved Faster RCNN augment the sensitivity. Through experimental comparison, the sensitivity of the baseline without a deconvolutional layer is 0.817, the sensitivity of the baseline with the anchor values of 4 is 0.895, and the sensitivity of the improved Faster RCNN increases to 0.946. The second phase of the system adopt a 3D RCNN to implement false positive detection, which means eliminating images that are actually negative but misjudged to be positive. Compared with 2D CNN, 3D CNN (Figure 6) can extract the 3D background features of candidates. Due to the 3D structure of CT images, 3D CNN can extract more features, and the experience demonstrates its sensitivity is higher than that of traditional 2D CNN.

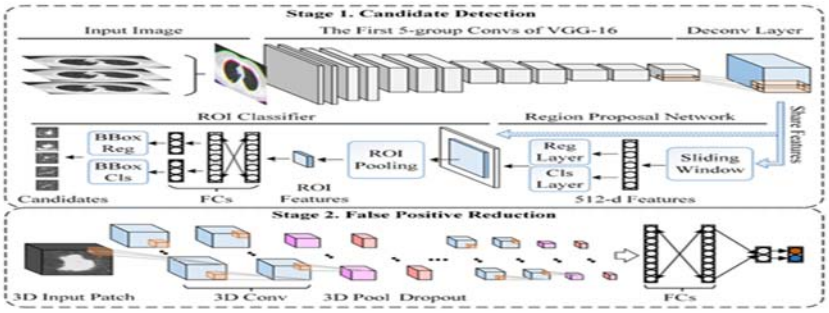

Fig 6. In the first phase, the last layer of shared CNN adds Deconv Layer.

In the second stage, the 3D CNN that is more adaptive to the CT image replaces the 2D CNN to complete the false positive reduction [12].

5. Conclusion

This paper introduced two common techniques in computer vision, image classification and object detection, and their application in non-medical technology. This paper took CNN and RCNN as

examples to analyze and discuss image classification and object detection, and compared the architecture characteristics of these two algorithms and their application scenarios in medical technology. By discussing the application of the two algorithms, it could be found that the CNN processing object is broad and extensible. Many traditional manual classification methods in medical image processing can be replaced by trained CNN; RCNN usually engages in object detection on medical images, such as detecting nodules, tumors, and cells infected with viruses. In more complicated application scenarios, the two algorithms are often integrated to process medical images such as CT, X-ray and smear. The development of computer vision in the medical field still has certain limitations, such as the scarcity of datasets, limited improvement of precision, and complex network architecture.

Referneces

[1]. LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11), 2278-2324.

[2]. Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 580-587).

[3]. Girshick, R. (2015). Fast r-cnn. In Proceedings of the IEEE international conference on computer vision (pp. 1440-1448).

[4]. Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. In Advances in neural information processing systems (pp. 91-99).

[5]. Rajpurkar, P., Irvin, J., Zhu, K., Yang, B., Mehta, H., Duan, T., & Lungren, M. P. (2017). Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning. arXiv preprint arXiv:1711.05225.

[6]. Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems (pp. 1097-1105).

[7]. Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.

[8]. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., & Rabinovich, A. (2015). Going deeper with convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1-9).

[9]. Szegedy, C., Ioffe, S., Vanhoucke, V., & Alemi, A. A. (2017, February). Inception-v4, inception-resnet and the impact of residual connections on learning. In AAAI (Vol. 4, p. 12).

[10]. He, K., Zhang, X., Ren, S., & Sun, J. (2014, September). Spatial pyramid pooling in deep convolutional networks for visual recognition. In European conference on computer vision (pp. 346-361). Springer, Cham.

[11]. Hung, J., & Carpenter, A. (2017, July). Applying Faster R-CNN for Object Detection on Malaria Images. In 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW) (pp. 808-813). IEEE.

[12]. Ding, J., Li, A., Hu, Z., & Wang, L. (2017, September). Accurate pulmonary nodule detection in computed tomography images using deep convolutional neural networks. In International Conference on Medical Image Computing and Computer-Assisted Intervention (pp. 559-567). Springer, Cham.