An Hybrid Parallel Network Structure for Image Classification

Muhan Li1,* and Yijian Jin2
1International Division, Wuhan Britain-China School, Wuhan, China
2Beijing Huijia Private School, Beijing, China

*Corresponding author email: mhan.li@outlook.com

Abstract. The image classification task is to divide certain image target into the specific class which they belong to. The deep neural network simplifies image classification by sending input image data into the neural network to learn feature extraction and having the output results automatically. Existing network structures have been able to achieve or exceed human capabilities in image classification tasks. However, these image classification technologies still have problems such as over-fitting network model, insufficient generalization abilities and long convergence time. Interference caused by complex environments may also cause the performance degradation. This paper proposes a hybrid parallel neural network structure for this kind of artificial intelligence task. By adopting the K-parallel connection CNN structure and using different scales of RNN/CNN structure mixed-layer modes, the effects of image’s size on different objects and different shooting angles can be reduced in image classification. Experiments on the datasets show that the proposed structure will have the advantage of accuracy and computational efficiency both in the validation datasets and the test sets.

Keywords: Hybrid parallel; CNN; Image classification.

1. Introduction

Image classification technology has long been used in the internet applications around the world. It is to divide certain object images into the categories which they belong to. The traditional image classification algorithms adopt the path of feature extraction and training classifier. Most of feature extraction adopts HOG/SURF/LBP and other easily distinguishable features. From around the year of 2010, the deep neural network has gradually become the mainstream. It is only necessary to input image data into the neural network, and the system is responsible for automatically learning feature extraction and outputting the results.

There are many competitions in this field, and computer vision practitioners actively participate. from the initial grayscale image handwriting digital recognition task to the iceberg image classification competition, Image Recognition Competition, and the largest competition ImageNet. These competitions give some benchmark models and compare them to new technologies.

Textures were originally used as important features of objects or regions of interest in photomicrographs, aerial photographs and satellite images [1], for example using easily calculated texture features based on grayscale spatial dependence. Such features are used in class identification tasks for a variety of different image data. From image to video in pixel form, many images are produced and viewed. How to find the right representation for these images and compare them is very difficult. Reference [2] use color histogram and the all-win method to complete the classification task.
Traditional feature extraction methods cannot improve the recognition of tasks such as handwritten numbers or traffic signs to humans. The deep artificial neural network architecture can accomplish these tasks well [3]. CNN had been proved that it is well suited for computer vision tasks. By mapping the receptive fields and creating large network depths for all neurons, a structure like the neural layers of the retina and visual cortex was created. In the MNIST handwritten benchmark test, this structure achieves results close to the level of human brain processing for the first time. It is twice as high as humans in terms of traffic sign recognition.

In the work of paper [4], a residual attention network was proposed which uses an attention mechanism that can be combined with the most advanced feedforward network manner. The analysis of the CIFAR-10 and CIFAR-100 datasets was performed to verify the effectiveness of each of the above modules. Both deep CNN and non-deep MLP are common neural network learning methods. Literature [5] proposed a MLP-CNN structure based on depth space and spectrum identification to combine the advantages of the two, and deal with some of the limitations of CNN. The MLP- is given on a given dataset. The effectiveness of the CNN classifier.

With the rapid iterative development of deep neural networks, there are constantly better technologies appearing in image classification techniques and applications, and the correct rate is constantly approaching and surpassing human capabilities. However, image classification technology is still likely to suffer performance degradation. Over-fitting in the network, insufficient model generalization ability, and long convergence time may cause good systems to fail to achieve the required recognition accuracy. In addition, due to interference caused by complex environments, such as light, occlusion, shooting angle, etc. In this paper, one image classification of hybrid parallel neural network is proposed. By layering the layers of convolutional neural networks and using a variety of neural networks, the image classification open source dataset is trained/tested. The difference from the similar network structure is that, this method has better precision and robust performance in the test data set. The organizational structure of the paper is: The second section includes the basic image classification technology and the situation of convolutional neural network; the third part details the structure of the hybrid parallel convolutional neural network; the fourth section shows the test results on the open source dataset and the last one is conclusion.

2. Image Classification and CNN Structure

2.1. Image Classification Technology

Recently, a zero shot deep learning tool has attracted researchers' attention. Attribute-based learning involves the introduction of an intermediate space called the attribute layer. Attributes correspond to advanced attributes of objects that are shared between multiple categories. It will be found by the machine or human. Each category can be represented as a vector with a correlation degree, which is directly related to the attribution of it. Each category provide association is usually binary.

![Figure 1. Much work in computer vision devoted to image embedding [6].](image)

The popular attribute-based prediction algorithms require a classifier for each attribute. In order to classify new images, the learned classifier is used to predict their attributes. The scores of attribution mixed to category grade scores. This two-step strategy is called Direct Attribute Prediction (DAP), which can be solved by using the tag embedding framework. For example, the attribute classifier is to overall strategy of DAP, which will be predicted. It is the best, but not necessarily the best when
predicting classes. The tag embedding framework, which is shown in Figure 1. Embedding each class into the space of the attribute vector. It uses a structured output learning form and introduces the metric compatibility function for the image and the tag, learning the parameters from tagged sample to keep in the case of a given image [6]. The identification includes how the category will be established with some advantage. The MugNet is proposed in the paper [7] to be applicable for the hyperspectral images, which is processing hyperspectral images for the less samples. Mugnet can be regarded as a relatively simple neural network algorithm because it does not have too many parameters.

2.2. CNN Structure

Human visual system can effectively identify and locate objects in clutter scenes, but it is still difficult to achieve this for machines. The concept of consensus in deep neural networks may be a breakthrough on this point [8]. Deep convolutional neural network (CNN) has received considerable praise in image processing, but there are still many difficulties in processing multi-label scenes. For example, CNN needs dependency information, and recurrent neural network (RNN) is just used to deal with this kind of problem [9]. CNN Pool (HCP) is proposed [10] and shown in Fig.2, which characteristics are that CNN is shared and the largest pool is aggregated.

![Figure 2. An illustration of the infrastructure of HCP [10].](image)

2.3. Hybrid Parallel Neural Network Structure

Although existing image classification methods have been more robust and performance than previous methods, the speed and recognition accuracy on large-scale image data can be improved. SRCNN [11] was proposed in twenty seventy, which is shown in Fig.3. To upgrade the size of S by inserting more data (its goal is to better reference the real annotation recovery image in the later stage), the objective function is:

1) Patch extraction and representation

\[
F_1(Y) = \max(0, W_1 * Y + B_1)
\]  

(1)

Where \(W_1\) and \(B_1\) are filters and biases separately.

![Figure 3. The structure of SRCNN model [11].](image)

2) Non-linear feature mapping

Non-linear mapping gives every n1-D vectors to another n2-D vector:

\[
F_2(Y) = \max(0, W_2 * Y + B_2)
\]  

(2)
Where the dimension of $W_2$ is $n_1*1*1*n_2$ and the dimension of $B_2$ is $n_2$.

3) Reconstruct
Reconstruction often is a C-layer:

$$F(Y) = W_3 * F_2(Y) + B_3$$  \hspace{1cm} (3)

Where the dimension of $W_3$ is $n_2*3*f_3*c$ and the dimension of $B_3$ is $c$.

Although the above scheme adopts new form of a parallel convolutional network, it is mainly used for reconstruction of individual images, performance on image classification tasks has not been given, and the computational cost still existed in the system. We propose a hybrid parallel network structure for this reason. K parallel convolutional network connection structure is adopted, and the modes of mixed layers of RNN and CNN of different scales are adopted to alleviate the problems of different target image scales and different photographing angles in image classification.

Figure 4. The hybrid parallel network structure.

As shown in Fig.4, each parallel network contains at least 2 RNN or CNN structures. The advantage of this network structure is that, However, the system can save more images of targets with different scales and extract more feature information in our structure by training the mixed layer, while taking advantage of RNN and CNN. Because CNN has dimensional constraints on the input image data, that is, the image size requirements are consistent, and RNN does not have such constraints; in addition, the RNN model introduces the context factor for data input, and CNN does not.

3. Experiments and Evaluation
To verify the advantage of proposed hybrid parallel deep neural network, we performed experiments under TensorFlow environment. The dataset is Kaggle dataset, which is provided by Kaggle contains 12,500 pictures of cats and dogs, all named after cat.<number>.jpg or dog.<number>.jpg, so they can be tagged according to file name. Get_files() is used to read the data set, tag the data set according to the file name, and return the image and label in the form of a list. Due to limited computing resources, we selected a total of 5000 images as training and 1000 samples as test. In our experiments, image preprocessing such as rotation or stretching is not used, and the raw data is directly used for input.

Figure 5. Average training and testing accuracy experimental results.
First, 5-layers CNN is selected as baseline on training data, the training accuracy on dog-cat image classification is 82%, which is the best result in all three network structures. Compared with the baseline system, the SRCNN structure and our hybrid parallel network structure have the weak training accuracy, but they almost have better validation accuracy and testing accuracy on both validation datasets and testing datasets. It can be observed that, the testing accuracy by hybrid parallel network structure reaches 75%. On the other hand, through comparison tests, it can be found that the computational cost of these three network structures is reduced in turn. Therefore, it can be considered that the hybrid parallel convolutional neural network structure has good recognition accuracy and efficiency at the same time compared with the two methods.

4. Conclusions

We propose a hybrid parallel CNN structure for image classification. The K-parallel convolutional network connection structure is adopted, and the mixed layer modes of RNN and CNN with different scales are adopted to alleviate the problems of different target image scales and different shooting angles in image classification. The training and testing experiments on image classification open source datasets show that this structure has advantage on accuracy and computational efficiency in the verification dataset and test dataset than the existing methods.

References

[1] Haralick R M, Shanmugam K. Textural features for image classification[J]. IEEE Transactions on systems, man, and cybernetics, 1973 (6): 610-621.
[2] Chapelle O, Haffner P, Vapnik V N. Support vector machines for histogram-based image classification[J]. IEEE transactions on Neural Networks, 1999, 10(5): 1055-1064.
[3] Cireşan D, Meier U, Schmidhuber J. Multi-column deep neural networks for image classification[J]. arXiv preprint arXiv:1202.2745, 2012.
[4] Wang F, Jiang M, Qian C, et al. Residual attention network for image classification[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017: 3156-3164.
[5] Zhang C, Pan X, Li H, et al. A hybrid MLP-CNN classifier for very fine resolution remotely sensed image classification[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2018, 140: 133-144.
[6] Akata Z, Perronnin F, Harchaoui Z, et al. Label-embedding for image classification[J]. IEEE transactions on pattern analysis and machine intelligence, 2016, 38(7): 1425-1438.
[7] Pan B, Shi Z, Xu X. MugNet: Deep learning for hyperspectral image classification using limited samples[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2018, 145: 108-119.
[8] Ciresan D C, Meier U, Masci J, et al. Flexible, high performance convolutional neural networks for image classification[C]. Twenty-Second International Joint Conference on Artificial Intelligence, 2011.
[9] Wang J, Yang Y, Mao J, et al. Cnn-rnn: A unified framework for multi-label image classification[C]. Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 2285-2294.
[10] Wei Y, Xia W, Lin M, et al. HCP: A flexible CNN framework for multi-label image classification[J]. IEEE transactions on pattern analysis and machine intelligence, 2016, 38(9): 1901-1907.
[11] Zeng K, Ding S, Jia W. Single image super-resolution using a polymorphic parallel CNN[J]. Applied Intelligence, 2019, 49(1): 292-300.