Scholar Publishing Group
International Journal of Neural Network
https://doi.org/10.38007/NN.2022.030308
ISSN 2790-2692 Vol. 3, Issue 3: 65-71

Product Information Classification based on Convolutional Neural Network

James Yong Liao*

Philippine Christian University, Philippine

James.liaoyong@pcu.edu.ph

**corresponding author*

Keywords: Convolutional Neural Network, VGG16 Model, Commodity Information, Commodity Classification

Abstract: With the growing e-commerce market and the increasing diversity of commodities, it brings great challenges for platforms and merchants to quickly and accurately annotate commodities. Multi-task commodity image classification technology combining attribute prediction, category classification and other tasks comes into being. This paper mainly studies the classification of commodity information based on convolutional neural network (CNN). Firstly, this paper analyzes the principle of CNN as the research basis, and uses the improved VGG16 CNN to build the structure of clothing product classification algorithm. Through the experimental results, we can know that the classification algorithm constructed in this paper can improve the goal of clothing image classification accuracy.

1. Introduction

With the continuous expansion of e-commerce platform, commodities are becoming more and more diversified, which brings great challenges to the platform and businesses to accurately label commodities, and also increases the difficulty for consumers to find commodities [1-2]. The e-commerce platform establishes a fine-grained hierarchical division system by organizing massive commodities in an orderly manner. On the one hand, it can easily recommend similar commodities for customers and obtain pleasant shopping experience [3-4]. The distribution of commodity attribute information will vary with different commodity categories. Some attribute information is specific to a commodity. For example, electronic products contain power attribute, but clothing products may not. On the concrete implementation, multitasking, deep learning method can be used to image category classification and properties prediction two tasks at the same time to study, learn two tasks associated with each other, using the properties prediction task auxiliary category classification task, to improve the performance of the category classification task, learning resources and reduce waste of [7-8].

In the field of image classification, compared with the research on natural images and individual images, less attention has been paid to commodity images. The driving force of commodity image classification is the rapidly developing e-commerce technology [9]. With the development of deep learning, more and more product classification models begin to adopt end-to-end CNN [10]. Some scholars have proposed Decaf model based on CNN to solve the problem of fine-grained commodity image recognition. The main idea of this model is transfer learning, which transfers the features of the activation layer and increases the recognition accuracy on the premise of reducing the amount of computation [11]. Some scholars built a CNN embedded with multiple spatial attention modules to improve the classification accuracy of goods by using predefined hierarchical division and attribute information of product images [12]. In the early stage of fine-grained image classification research, the main research is based on artificial annotation information methods, such as the salient space discriminant method [13] to find discriminative regions by analyzing image semantic features. Later, researchers began to apply some more expressive features, such as Fisher vector coding, POOF, SIFT, etc. to fine-grained image classification, which could improve the classification accuracy to about 50%-60% [14]. However, due to the limited feature description ability, the classification effect of the method based on manual annotation information is not very ideal.

Product image classification methods can provide technical support for product search, product positioning strategy formulation, intelligent product recommendation, product classification error correction, etc. Product image classification plays an important role in the development of e-commerce platforms.

2. Product Classification Algorithm Based on CNN

2.1. CNN

CNN can extract multi-level features of images, and can basically comprehensively express all semantic features of target objects, so as to carry out detection and recognition tasks more accurately [15]. CNN is to make the sharing of neural network weights in different position, has a certain translation, scaling and tilt invariance, CNNs of neurons with convolution kernels only within the scope of sliding neurons respond to each other, make its have the sparse connectivity and the sharing of the weights of features, Thus, it has excellent performance ability in the field of large-scale image processing [16-17].

CNN mainly includes input layer for input data, convolutional layer Cx for convolutional operation, pooling layer SX for downsampling, fully connected layer and output layer using activation function. The working process of CNN is that firstly, the input image is convolved and biased by the convolutional layer to get the feature map. Secondly, the pooling layer selects the feature map output by the previous layer to remove the redundant features and reconstruct the new feature map. Finally, it is fed into the classification function through FC layer for recognition, so as to complete the object recognition work [18].

CNN is mainly used for local connection and weight sharing through convolution kernel. Convolution is a method to extract local features of image and reduce feature parameters.

Convolution operation is the process of the fixed size convolution kernels as a sliding window on the input figure in the order from left to right, from top to bottom to slide, all the pixels are covered at least once, convolution kernels moves to the right and down the sliding step length are l pixels, for input sliding window of feature points on the drawing or pixels, First, it multiplicates the feature points or pixels in the convolution kernel point-to-point, and then obtains the convolution result through the accumulation operation of the product results. When the convolution kernel and the step size are both greater than L, zero-complement operation is performed on the input graph as the case may be.

The pooling layer is also called the downsampling layer. Different from the convolution layer, the pooling layer mainly operates on the non-coincident areas in the image and compresses the input feature map. The pooling layer reduces the number of neurons, the connection after the convolution layer reduces the data dimension, avoids the overfitting problem that occurs when the convolution results are directly input into the classifier, and makes the features robust to translation, rotation and other operations.

Random pooling, average pooling, and maximum pooling are common pooling operations. The pooling operation process is similar to the convolution operation process, which is to slide the sliding window of fixed size on the input feature map from left to right and from top to bottom, while all pixels are covered only once, and the step length of right and down is generally the size of the sliding window. For the element values in the sliding window, the pooled value of the region is calculated according to the operation of maximum pooling and average pooling. Of these, maximum pooling is the most common pooling operation.

2.2. Based on VGG16 Classification Algorithm

In this paper, the basic VGG16 CNN framework is adopted, and the traditional CNN image classification idea is used to classify clothing images.

(1) Image feature extraction

Aiming at the phenomenon of model degradation, a deep residual network model is proposed, and residual modules are added to the original network structure. In this way, the incomplete degree of image information and semantic information will be reduced in the process of deepening the network level, and the deep network can achieve higher classification accuracy in the multi-classification task of clothing image.

The residual network structure uses the idea of cross-layer link and combines the original network structure layer with the identity mapping layer. When the network has reached a good training state, the identity mapping layer will be used in the network structure to avoid the occurrence of accuracy decline to the greatest extent. In the residual network structure, by using the method of "shortcut connection", the input X is directly mapped to the output, and the output result $H(X)=F(X)+X$ is obtained by corresponding operation with the output result $F(X)$ of the original network. Therefore, the final learning objective of the network is indirectly transformed into the learning residual $F(X)=H(X) -x$. When the learning result approaches 0, the accuracy of the target task will not decrease with the gradual deepening of the network structure. This network structure provides a new idea and direction for deepening the network level without reducing the accuracy of the task.

(2) Position of key marker points

In the process of predicting the position of key marker points in clothing images, this paper firstly uses multiple convolution cores to extract features from the original clothing images, aiming to generate high-resolution feature maps of key marker points in clothing.

The details are as follows: Firstly, a convolution kernel with the size of 1×1 is used to transform the dimension of the feature map of the clothing image to $28\times 28\times 64$, and the dimension of the feature map to $56\times 56\times 64$.

Then, the same two convolution operations with a convolution kernel size of 3×3 and one deconvolution operation with a convolution kernel size of 4×4 were used to transform the feature map dimension to $224\times 224\times 16$. Finally, a convolution kernel with the size of 3×3 and a convolution kernel with the size of 1×1 were successively used for convolution operation to generate a feature map of key clothing marker points with the dimension of $224\times 224\times 8$. By transforming the feature map size into $28\times 28\times 1$, the feature map was fused with the original clothing image feature map. Then complete the subsequent task of clothing image classification.

(3) Classifier

This paper uses cnn model and cross entropy loss function to classify clothing images in the process of network design.

The function of cross entropy in cnn structure is to describe the distance between the actual image output rate and the label output rate. The cross entropy value represents the distance between the two. In the garment image classification task based on CNN:

$$H(p,q) = -\sum p(x)\log q(x) \tag{1}$$

The formula of p and q represent clothing classification real value and clothing classified forecast, after cross entropy loss function to calculate the numerical may not meet the conditions and significance of probability distribution, so I need the final after softmax activation function, will become the form of probability distribution, data processing of clothing image classification task requires more of the algorithm. The significance of Softmax activation function is that it does not directly determine the unique classification result, but assigns corresponding probability values to the classification results of each classification output, the value is between 0 and 1, which is used to represent the possibility of each classification result.

$$Soft\max(Z_i) = \frac{e^{zi}}{\sum_{c=1}^{C} e^{zc}} \tag{2}$$

3. Algorithm Simulation Experiment

The data in this paper are mainly from Taobao, Jingdong, Amazon and other major shopping websites. There are 50 images of each type of goods, of which 10 are used as test samples and 40 are used as training samples. The images are from different sizes of the network, and the pixels are adjusted for the convenience of processing and normalization.

Data preprocessing stage to do the processing of the mean, so it is necessary to obtain the mean of the image file. The mean value file can be obtained by taking the mean value of all the training samples, that is, a mean value file in mean.binaryproto format can be generated. In the process of image preprocessing, the mean value is subtracted, and then these data are input into the network for training and testing, which can improve the training accuracy and speed to a certain extent.

4. Analysis of Experimental Results

4.1. Experimental Results of Different Models

Table 1. VGG16 network classification result

	10	20	30	40	50
Test Loss	1.42	0.85	0.46	0.43	0.37
Train Loss	1.38	0.79	0.31	0.41	0.31
Accuracy	18%	71%	83%	87%	92%

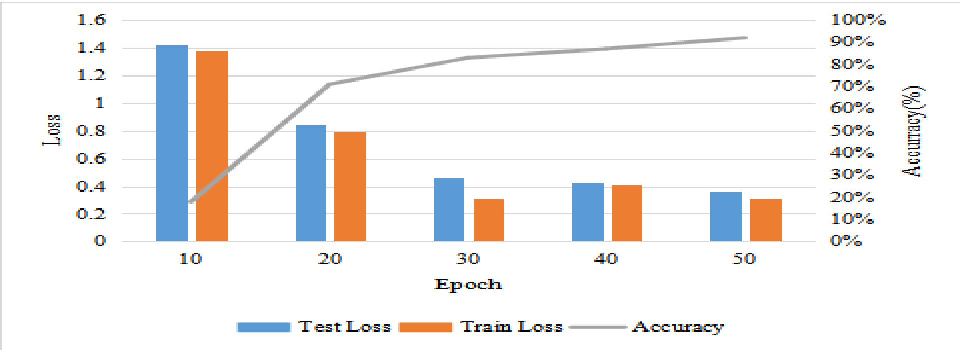

Figure 1. VGG16 network classification experimental data

The experimental results in this paper are the average values obtained from 10 experiments, and the experimental results after 50 epochs are shown in Table 1 and Figure 1. In this paper, the average classification accuracy of the network structure for clothing products is 90.7%, and the classification accuracy gradually increases with the increase of epochs. After 30 epochs, the classification accuracy of the experiment becomes stable. The loss function in the CNN is used to measure the difference between the real value and the measured value. The higher the convergence degree of the loss rate, the smaller the loss and the better the experimental effect.

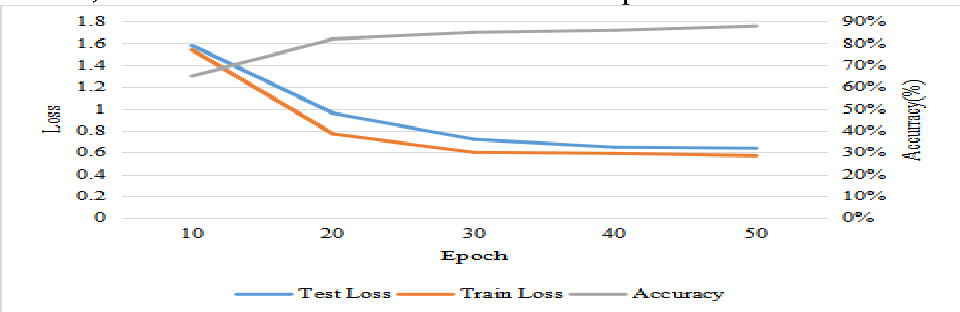

Figure 2. AlexNet classification results

As a comparison, we used the classical network structure AlexNet to test the experimental data in this chapter as well. FIG. 2 shows the evaluation results obtained from 10 experiments. After 50 epochs, the classification accuracy of AlexNet for this article reaches 87.5%. It can be seen from the classification experiment results that the network proposed in this paper achieves better classification results, and the average classification accuracy of AlexNet is improved by 3.2 percentage points in the classification of commodity data. It can also be seen that the improved network has faster convergence speed and lower loss rate. The experiment in this paper proves that the improved network has better network performance than AlexNet on the problem of commodity classification.

4.2. Improved Network Analysis

Table 2. Capacity of convolutional kernel in improved network

	Conv1	Conv2	Conv3	Conv4	Conv5
Receptive field size	6	18	59	110	142
Convolution kernel capacity	1	0.541	0.824	0.737	0.583

Table 3. Capacity of convolutional kernel in AlexNet

	Conv1	Conv2	Conv3	Conv4	Conv5
Receptive field size	12	54	98	132	165
Convolution kernel capacity	1	0.287	0.258	0.213	0.171

Table 2 and Table 3 show the receptive field university and convolution kernel capacity of each convolution layer of the network results of the proposed network and AlexNet. It can be seen that the convolution kernel capacity of the improved network in this paper is generally higher than that of AlexNet.

5. Conclusion

With the rapid development of multi-task deep learning in the field of computer vision, this paper mainly studies the multi-task deep learning model and its application in product image classification, aiming at the problem that commodity attribute information is not fully utilized in product image classification. The method used in this paper is to add the key marker prediction branch network into the structure of VGG16 CNN to enhance the accuracy of clothing category prediction. On this basis, two improvement directions, feature fusion and attention mechanism, are proposed to improve the accuracy of clothing image classification by combining the task of clothing image classification.

Funding

This article is not supported by any foundation.

Data Availability

Data sharing is not applicable to this article as no new data were created or analysed in this study.

Conflict of Interest

The author states that this article has no conflict of interest.

References

[1] Jain E. CNN Based Advertisement Classification Models for Online English Newspapers. Turkish Journal of Computer and Mathematics Education (TURCOMAT), 2021, 12(2):1687-1698. https://doi.org/10.17762/turcomat.v12i2.1505

[2] Wambugu N, Chen Y, Xiao Z, et al. A hybrid deep CNN for accurate land cover classification. International Journal of Applied Earth Observation and Geoinformation, 2021, 103(4):102515. https://doi.org/10.1016/j.jag.2021.102515

[3] Shingo, Mabu, Yoshiaki, et al. Landslide Classification from Synthetic Aperture Radar Images Using CNN with Multichannel Information. Journal of Signal Processing, 2020, 24(2):61-73. https://doi.org/10.2299/jsp.24.61

[4] Babatunde R S, Adigun O, Folasade O. An Assessment of Gpu and Cpu based CNN for Classification of White Blood Cells. ANS. Advances in nursing science, 2020, 9(7):554-559. https://doi.org/10.35940/ijitee.F4068.059720

[5] Shim J, Kang S, Cho S. Active Learning of CNN for Cost-Effective Wafer Map Pattern Classification. IEEE Transactions on Semiconductor Manufacturing, 2020, PP(99):1-1.

[6] Lee S Y, Lee S K. Deep CNN with new training method and transfer learning for structural fault classification of vehicle instrument panel structure. Journal of Mechanical Science and Technology, 2020, 34(11):4489-4498. https://doi.org/10.1007/s12206-020-1009-3

[7] Hormese J, Chandran S. A CNN Approach to Road Classification from Satellite Images. Journal of Theoretical and Applied Information Technology, 2018, 96(23):7917-7927.

[8] Scott G J, England M R, Starms W A, et al. Training Deep CNNs for Land–Cover Classification of High-Resolution Imagery. IEEE Geoscience and Remote Sensing Letters, 2017, 14(4):549-553. https://doi.org/10.1109/LGRS.2017.2657778

[9] Widhiarso W, Yohannes, Prakarsah C. Brain Tumor Classification Using Gray Level Co-occurrence Matrix and CNN. IJEIS (Indonesian Journal of Electronics and Instrumentation Systems), 2018, 8(2):179-190. https://doi.org/10.22146/ijeis.34713

[10] Kumar N, Kaur N, Gupta D. Red Green Blue Depth Image Classification Using Pre-Trained Deep CNN. Pattern Recognition and Image Analysis, 2020, 30(3):382-390. https://doi.org/10.1134/S1054661820030153

[11] Tfekc P, Bekta M. Author and genre identification of Turkish news texts using deep learning algorithms. Sādhanā, 2022, 47(4):1-10. https://doi.org/10.1007/s12046-022-01975-3

[12] Ben-Yair I, Ben Shalom G, Eliasof M, et al. Quantized CNNs through the lens of partial differential equations. Research in the Mathematical Sciences, 2022, 9(4):1-21. https://doi.org/10.1007/s40687-022-00354-y

[13] Kruthiventi S, Ayush K, Babu R V. DeepFix: A Fully CNN for Predicting Human Eye Fixations. IEEE Transactions on Image Processing, 2017, 26(9):4446-4456. https://doi.org/10.1109/TIP.2017.2710620

[14] Acharya U R, Fujita H, Lih O S, et al. Automated detection of arrhythmias using different intervals of tachycardia ECG segments with CNN. Information Sciences, 2017, 405:81-90. https://doi.org/10.1016/j.ins.2017.04.012

[15] Acharya U R, Fujita H, Lih O S, et al. Automated detection of coronary artery disease using different durations of ECG segments with CNN. Knowledge-Based Systems, 2017, 132(sep.15):62-71. https://doi.org/10.1016/j.knosys.2017.06.003

[16] Paoletti M E, Haut J M, Plaza J, et al. A new deep CNN for fast hyperspectral image classification. Isprs Journal of Photogrammetry & Remote Sensing, 2017, 145PA(NOV.):120-147. https://doi.org/10.1016/j.isprsjprs.2017.11.021

[17] Evo I, Avramovi A. CNN Based Automatic Object Detection on Aerial Images. IEEE Geoscience & Remote Sensing Letters, 2017, 13(5):740-744. https://doi.org/10.1109/LGRS.2016.2542358

[18] Haenssle H A, Fink C, Schneiderbauer R, et al. Man against machine: diagnostic performance of a deep learning CNN for dermoscopic melanoma recognition in comparison to 58 dermatologists. Annals of Oncology, 2018, 29(8):1836-1842.