Research on E-Commerce Platform Product Image Recognition Based on ResNet Network

Kunxian Wang
School of Computer Science, China University of Geosciences, Wuhan, China

Keywords: Convolutional Neural Network, Residual Network, E-Commerce

Abstract: With the development and popularity of the Internet and smart mobile devices, online shopping has gradually become the main way for contemporary people to shop, the massive amount of commodity information makes it difficult for users to choose, how to get the correct commodity information has become a difficult problem, and manual labelling of commodity categories is very inefficient. How to improve the correct rate of commodity classification, numerous academics have conducted extensive research in this field., the current mainstream method is to use convolutional neural networks for commodity picture classification. In this paper, we carry out a study on the classification of commodity pictures on e-commerce platforms by constructing three ResNet networks with different depths, ResNet18, ResNet34 and ResNet50, respectively, and explore the practical significance by training and comparing the classification accuracy by using the commodity pictures directly downloaded from the famous e-commerce platforms in China. From the experimental results, evidently, as the depth of network layers increases, the performance results of the network are getting better and better, while all three networks have achieved a high accuracy rate, which indicates that convolutional neural networks have application value in the classification of commodity pictures.

1 INTRODUCTION

Amidst the swift evolution of the Internet and the pervasive adoption of electronic devices such as smartphones and portable computers, individuals' proclivity to partake in online shopping is discernible. Over the past few years, a marked surge has been observed in the diversification of merchandise accessible for online procurement. As a result, the challenge of swiftly and accurately locating the desired products among this vast array of information has become an urgent matter to address (Wu 2021). Unlike traditional offline shopping methods, online shoppers are unable to physically interact with products. Instead, they rely on the information provided by sellers, including product images and accompanying textual details. As e-commerce continues to grow, the volume of product information has exploded. Hence, there is a pressing need for automated product classification to enable users to efficiently locate their desired items, ultimately enhancing overall efficiency (Lowe 2004).

Currently, the most widely used method is the automatic categorization of commodities based on text keywords.

Merchants need to provide text information to the commodities when they are on the shelves, and the e-commerce platform completes the classification of commodities according to the category information provided by the merchants. When users enter keywords to search, the e-commerce platform can intelligently match the text keywords and recommend related products. This method is simple to operate and fast to query, but it requires merchants to provide accurate commodity information, and this over-reliance can easily lead to misclassification (Kai-Qi et al 2014).

Commodity images can provide most of the commodity information in the most intuitive way, and if e-commerce platforms can use the classification method based on commodity images, it can bring a better user experience to both merchants and users. Researchers in various countries began to use traditional machine learning methods for merchandise image classification many years ago, but because merchandise images are more complex than

Wang, K.
Research on E-Commerce Platform Product Image Recognition Based on ResNet Network.
DOI: 10.5220/0012804000003885

In *Proceedings of the 1st International Conference on Data Analysis and Machine Learning (DAML 2023)*, pages 87-91
ISBN: 978-989-758-705-4

general dataset images, it is more difficult to extract features. With the development of deep learning, researchers started building various network models for image classification. Convolutional neural network is a representative network, which has been widely used in the field of image classification and achieved better results. In this paper, we will use the ResNet network in convolutional neural network for the research of image classification (He et al 2023).

2 METHOD

The experimental method used in this paper is to obtain image data on the Internet and use ResNet(Residual Neural Network) network model for training to obtain a model that can classify commodity pictures. ResNet has a good performance in degradation problems and gradient disappearance problems, which is the main reason for choosing this model in this paper. Further details will be provided in the following explanation.

Conventional convolutional neural networks are confronted with a formidable predicament. As the quantity of network layers is augmented, the network's capacity for intricate feature pattern extraction is enhanced, theoretically leading to potentially superior outcomes with increased model depth. Deep networks have been found to suffer from degradation problems, with network accuracy saturating or even decreasing when the network depth increases. The experimental results reveal that the 56-layer network is not as effective as the 20-layer network. This phenomenon does not stem from overfitting, given that the disparity between predicted and actual values during 56-layer network's training phase also manifests conspicuously.

Training deep networks poses challenges due to the problem of degeneracy. In the context of a shallow network, the introduction of additional layers in an upward manner to formulate a deep network may yield an extreme scenario. In this scenario, these supplementary strata may remain uninformative, effectively duplicating the shallow network's features. This implies that the newfound layers amount to an Identity mapping. In such an instance, the deep network is expected to exhibit performance on par with the shallow network, and degradation should not manifest.

Residual learning was introduced by Kaiming He's team to address the issue of degradation. In a stacked layer configuration, denoted as H(x), where x is the input, the objective is for the model to learn the residual, as in:

$$F(x)=H(x)-x. \quad (1)$$

Hence, the original acquired feature can be represented as in:

$$F(x)+x. \quad (2)$$

The rationale behind this lies in the fact that learning residuals is simpler compared to directly learning the original features. When the residual is eliminated, the stacking layer's function is limited to a constant mapping. This restriction prevents any degradation in network performance at a minimum. Nonetheless, it's crucial to acknowledge that the residual is rarely, exactly 0. The stacked layers can acquire novel features from input characteristics due to the presence of non-zero residuals, which leads to improved overall performance. Figure 1 illustrates the structure of residual learning (He et al 2023). This bears a certain resemblance to a 'short circuit' within an electrical circuit, thus constituting a shortcut connection.

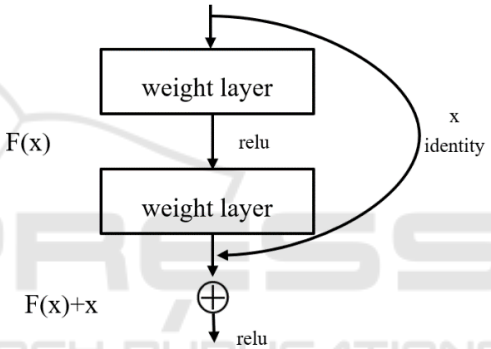

Figure 1: Residual learning unit.

In ResNet, one of the residual modules shown in Figure 1 is called Bottleneck. ResNet is available in different versions with different numbers of network layers such as 18, 34, 50, 101, and 152 layers and the structure of different layers is shown in Table 1 (He et al 2023).

3 EXPERIMENTAL METHODS

3.1 Experimental Framework

The operating system of this experiment is Windows 10, and the graphics card is GTX-1660ti. there are many mature open-source frameworks supporting convolutional neural networks and ResNet residual networks, such as TensorFlow, PyTorch, Keras, Caffe, and so on (Abadi et al 2016, Paszke et al 2023 & Jia et al 2014). Pythorch is chosen for this experiment, and the experimental process and result analysis are based on the Pytorch framework.

Table 1: Resnet Structure of Different Layers.

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

3.2 Data Description

The data used in this paper comes from Taobao, a Chinese e-commerce platform, where "Phone", "Shoes", "Men's clothing" and "Dress" are selected for product image downloads. There are about 1000 images in each category. The actual number fluctuates slightly due to the limitation of downloading images for some products. 80% of all images are used for training and 20% for testing. Part of the data is shown in Figure 2.

Since the image size is not the same, use resize to scale, after the end of processing each image size is 224 * 224.

Figure 2: Partial data presentation.

3.3 Model Construction

The ResNet network part of this experiment was built by directly calling the library in Pytorch, and ResNet18, ResNet34, and ResNet50 were used for comparison experiments. In addition to the network part, the experimentation also incorporates the utilization of the Adam optimizer. This optimizer amalgamates the merits inherent in both the AdaGrad and RMSProp optimizers. It adeptly facilitates the adaptive modulation of the learning rate, thereby expediting the convergence velocity (Kingma and Ba 2017, Duchi et al 2011 & Tieleman and Hinton 2012). For the loss function, the CrossEntropyLoss function was chosen, which is a loss function used to calculate the cross-entropy loss, which takes the labeled data and the probability distribution of the model output as inputs and calculates a scalar loss based on the labeled data and the model's prediction results. The experimental parameters are presented in Table 2.

Table 2: Experimental Parameters.

Learning rate	Batch_size	Epochs	Device
1e-4	64	30	GPU

4 EXPERIMENTAL RESULTS AND ANALYSIS

The performance of ResNet18, ResNet34 and ResNet50 with three different depths of residual networks on the same dataset is shown in Figure 3,

which shows that as the network's layer count increases, the correctness rate is incrementally increasing from 85.90% to 88.45%, a more obvious improvement, which can be reflected in the fact that the ResNet network successfully solves the problem of degradation that occurs with the deepening of the network's layers. This observation further demonstrates the efficacy of the ResNet architecture in mitigating the degradation phenomenon that manifests as the network's layer count increases.

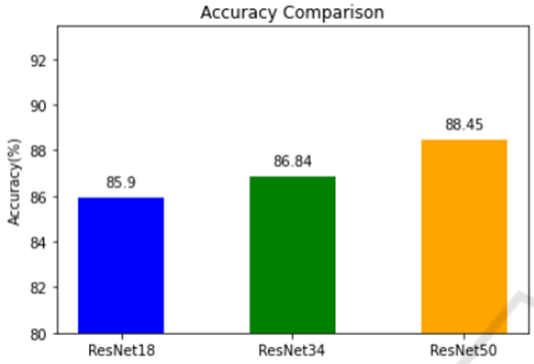

Figure 3: Comparison of ResNet correctness for three depths.

Figure 4: Partial results.

After outputting the prediction results of the test set part of the results are shown in Fig. 4. Observation of the prediction results can be found, for cell phones and shoes such as a large difference between the goods, basically, all predicted correctly, but men's clothing and dresses such as the same clothing category of goods will be more confusing, it may be that some skirts are more neutral or some men's clothing is more ornate. This result indicates that the ResNet network has practical application for the classification of goods on the e-commerce platform.

5 CONCLUDING REMARKS

This paper investigates the practical utility of automated product classification for e-commerce platforms, where ResNet networks of varying depths were constructed for experimental purposes, and their performances were compared. The image dataset used for experimentation was sourced from prominent e-commerce shopping websites in China. An important advantage of the methodology employed in this study is the elimination of the need for a priori manual feature extraction, with the models utilized yielding notably high classification accuracy.

One notable merit of incorporating the ResNet architecture in this research lies in its inherent capability to address network degradation and the issue of gradient vanishing, both common challenges encountered when training deep neural networks. The design of ResNet, featuring shortcut connections, facilitates a smooth flow of gradients during the backpropagation process, thus rendering the training process more robust and efficient. This attribute holds particular significance when dealing with intricate datasets and deep network architectures.

Nevertheless, despite the evident strengths of ResNet, some limitations persist in the design of this article, such as occasional misclassifications of similar product categories. While this aspect may not be solely attributed to the architectural choice but may benefit from further refinement through the fine-tuning of product categories and the exploration of advanced techniques for enhancing user-friendly search optimization. In the foreseeable future, our research endeavors aim to delve more profoundly into harnessing the complete potential of ResNet while conducting further investigations and refinements to produce more resilient product classification models.

REFERENCES

W. X. TUO Wu, “Fashion Style Recognition Based on Improved Edge Detection Algorithm”, Journal of Textile Research, vol. 42, no. 10, pp. 157−162, Oct. 2021.

D. G. Lowe, “Distinctive Image Features from Scale-Invariant Keypoints,” International Journal of

Computer Vision, vol. 60, no. 2, pp. 91–110, Nov. 2004.

H. Kai-Qi, R. Wei-Qiang, T. Tie-Niu, "A Comprehensive Survey of Image Object Classification and Detection Algorithms", Journal of Computer Science and Technology, vol. 37, no. 6, pp. 1225−1240, 2014.

K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778. Accessed: Aug. 23, 2023.

M. Abadi et al., "TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems." arXiv, Mar. 16, 2016.

A. Paszke et al., "PyTorch: An Imperative Style, High-Performance Deep Learning Library," in Advances in Neural Information Processing Systems, Curran Associates, Inc., 2019. Accessed: Aug. 24, 2023.

Y. Jia et al., "Caffe: Convolutional Architecture for Fast Feature Embedding," in Proceedings of the 22nd ACM international conference on Multimedia, in MM '14. New York, NY, USA: Association for Computing Machinery, Nov. 2014, pp. 675–678.

D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization." arXiv, Jan. 29, 2017.

J. Duchi, E. Hazan, and Y. Singer, "Adaptive subgradient methods for online learning and stochastic optimization.," Journal of machine learning research, vol. 12, no. 7, pp. 257–269, 2011.

T. Tieleman and G. Hinton, "RMSProp: Divide the Gradient by a Running Average of its Recent Magnitude," University of Toronto, Toronto, 2012.

G. E. Hinton, "Connectionist learning procedures," Artificial Intelligence, vol. 40, no. 1–3, pp. 185–234, Sep. 1989.