

Deep Learning based Model for Prohibited Goods Detection

Zhao Liu¹, Ying Ruan², Chun Chen³

^{1,2,3}38 Zheda Road · Hangzhou · Zhejiang Province · 310027 · P. R. China

liuzhao@zju.edu.cn, yruan90@163.com, chenc@zju.edu.cn

Keywords: Prohibited goods; Deep Convolutional Neural Networks; Deep Belief Networks

Abstract With the rapid development of Internet technology, massive images are used on e-commerce sites to show product details. Among those images there are some of contraband, which seriously harm the social harmony, thus it is important to automatically identify them. In this paper, we propose an effective method aims at solving this problem, in contrast with traditional methods, we do not use human designed visual features and classification models, but combine deep features and deep learning model. Experiment results show that our method outperform previous human designed features and visual models tremendously.

Introduction

The rapid development of e-commerce make it possible to trade prohibited goods, such as guns and bombs on Internet. It is time cost to search for humans to label all the prohibited goods in big e-commerce website such as Amazon and Taobao.

Most of the existing methods for prohibited goods detection are based on text analysis. That is, they extract the text information from the images, depart the text information to several individual words, and recognize the prohibited words by using various classification models. Although they can get good results, in real applications, the prohibited goods are always unlabeled. Thus it is essential to solve the problem of automatically detecting prohibited goods.

Deep learning have been paid much attention since the year of 2006, the original ideal of deep learning is to simulate the hierarchical transferring process in human neural systems. In recent years, deep learning was applied to various field, a typical example is the voice recognition system proposed by MicroSoft [1][2], by using deep learning technology, the recognition errors was successfully reduced by 25 percent.

In this paper, we proposed a deep learning based model for detecting the prohibited goods in images, by using the deep convolutional neural networks (CNN), we get representative features which are more adaptive to human visual system, then we use the deep features as input to train the deep belief networks (DBN), and learn a new distribution on the training samples. In order to improve the detection performance, we further use the selective search window to allocate the prohibited goods.

Related Work

Feature Extraction Traditional human designed features includes the color based features, the texture based features and the shape based features. Smith and Chang [3] Proposed the color sets feature, it is much faster than color histogram but keep near the same effect. Color moment was proposed in 1995, it is easy to implement, however, when used in indexing, it is inefficient. The most used texture based feature is the Gray-level Co-occurrence Matrix (GLCM) proposed by Kreyszig et al.[4] Also, the model based methods, such as [5] and [6], are widely used for extracting the texture features. The shape based features are divided into region features and contour features, region features refer to the perimeters, area, or the longest axis of objects, while computation of contour features is always through Hough Transform[7].

Image Classification. Decision tree has a typical tree form structure, the classification process of decision tree includes feature selecting, tree constructing and edge cutting. Compared with other classification methods, it is able to classify huge data in relatively short time, but neglects the

relation between different features. KNN is easy to implement, however, as a lazy learning method, when the input samples lose the balance, it will have significant impact on the final results. SVM is a binary classification model, when applied to non-linear data samples, nonlinear kernel function should be used. SVM is able to solve the problems when input samples are not sufficient, however, as the kernel function of SVM is human designed, it lacks initiative when applied to non-linear data.

PROPOSED APPROACH

Deep Feature Extraction In the year of 2012, Hilton et al.[8] use their own designed deep Convolutional Neural Networks (CNN) in ImageNet Challenge and obtain excellent results, their model is show in Figure 1. It contains five convolution layers, two full-connected layers and one output layer of dimension 4096. The authors use stochastic gradient descent and back propagation algorithms in the training stage, and a rectified linear unit as output, this model is able to accurately classify various object types.

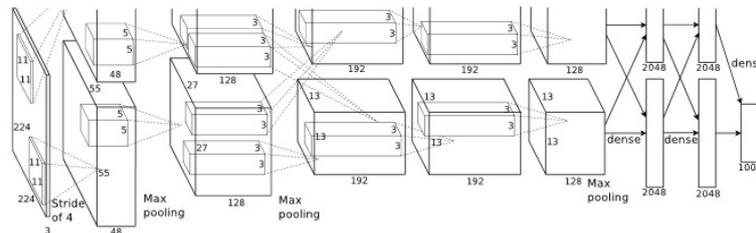


Fig. 1 The deep convolutional neural networks.

As we have massive of images, labelling all of them is a time cost work, thus we apply transfer learning [10] in the training process to improve the efficiency. Following the ideal of transfer learning, we use the pre-trained deep networks on ImageNet as the feature extractor and map the raw images into the feature space. In details, we first normalize all images to a unique size of $224 \times 224 \times 3$, then we input the images into the pre-trained network, after the computation process, the 4096 dimension vectors output of the first and second fully-connected layers, are used as the final features.



Fig. 2 The five highest scored windows predicted by using Selective Search method.

Detection for Prohibited Goods We have introduced the most used methods for classification in Section II, in order to improve the classification accuracy, we use Deep Belief Network (DBN) [11] for classification and detection. DBN was proposed by Hinton et al. in 2006, it is constructed by multiple Restricted Boltzmann Machines, each of which can be seen as an individual cluster and can represent a two-layer neural network. DBN can be applied both to features extraction and classification.

For detecting the prohibited goods, we first divide the input images to the training and testing sets, for the training images, we train DBN based on the features generated by CNN. In order to improve the detection accuracy, in the test stage, we select several candidate windows in each test image and use the pre-trained DBN to classify which window contains the target prohibited goods. Sliding window method is a common used method for extracting the candidate windows in previous work, however, when applied to classification, the sliding window method lacks the object information and is time consuming. In replace, we employ Sande et al.'s [12] Selective Search window to generate the candidate windows. Selective Search add an preprocessing segmentation stage, which is able to reduce the window numbers and focuses more on the object-like regions. An example of the windows generated by Selective Search method is illustrated in Figure 2.

Experiments

Dataset There exists no specified prohibited goods image dataset before, so we created the “Dangerous” database by ourselves, all the images are collected from the e-commerce websites, this database is constructed by 10,000 positive images and 10,000 negative images. The prohibited goods are divided into four subclasses, including guns, knives, bullets and bombs.

TABLE 1 BINARY LABEL CLASSIFICATION RESULTS OF DIFFERENT FEATURES

Input Features	Dimension	Classification Accuracy
Gray-Value Feature	784	74.25% (2970/4000)
Color Feature	968	85.30% (3412/4000)
Gist Feature	512	82.12% (3285/4000)
Deep Feature	4096	90.40% (3616/4000)

Results for Binary and Multi label Classification For evaluating the classification accuracy of different features, we use them as the input of SVM classifier, we use 80% of the all images for training and 20% for testing. The binary classification results of different features are shown in Table I. From the result we can see that by using deep feature we can get significant improvement for our prohibited images, moreover, the additional training time cost for SVM is negligible.

We also perform experiments to compare different classifiers, including SVM, KNN and Decision Tree. The results are illustrated in Table III, from the results we can see that by using Deep Belief Network as the classifier our method is able to outperform all the traditional classifiers, which further certify the effectiveness of deep learning based methods on classification.

TABLE 2 BINARY LABEL CLASSIFICATION RESULTS OF DIFFERENT CLASSIFIERS

Classifiers Used	Classification Accuracy
SVM	47.73% (525/1100)
KNN	57.64% (634/1100)
Dicision Tree	58.82% (647/1100)
DBN	69.94% (766/1100)

Results for Prohibited Goods Detection Results for the detection of different prohibited goods are illustrated in Table IV, in the detection stage, we use the same settings—CNN for image feature extraction and DBN for classification. From the results we can see that our method get better results on “Guns” and “Knives” images than “Bullets” and “Bombs”. It is due to the fact that the training samples for Bullets and Bombs are much less, which cause insufficient learning for the deep model.

TABLE 3 CLASSIFICATION RESULTS FOR DIFFERENT PROHIBITED GOODS

Classifiers Used	Classification Accuracy
SVM	47.73% (525/1100)
KNN	57.64% (634/1100)
Dicision Tree	58.82% (647/1100)
DBN	69.94% (766/1100)

Finally, we trained various DBN models under different model weights, iteration times and learning ratios, results are shown in Figure 3. Number in x-axis represents different iterations times, y-axis represents the reconstruction error. And the curves of different color represent different parameters used. From the results we can see that the reconstruction error decreases as we use more

iteration times. In addition, when the initial weight is set to 0.0002 and learning weight is set to 0.1, our method performs best.

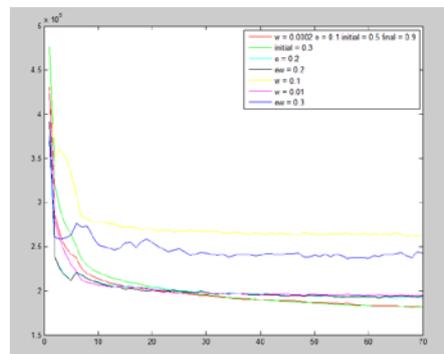


Fig. 4 Reconstruction errors under different weights, iteration times and learning ratios.

Conclusion

In this paper we propose a novel deep learning based method for detecting the prohibited goods in image. Different from previous human designed visual features, we use deep visual features trained on the deep convolutional neural networks as the input of our model, we retrain the deep belief networks based on the deep features detect the prohibited goods. Experiments on our own collected prohibited goods dataset have shown that our method can largely improve the accuracies of classification and detection. In the future, we will pay more attention to incremental learning for our model, and reducing time cost, in addition, we intend to apply our method to more other problems.

Acknowledgement

This work was supported by the National High Technology Research and Development Program of China(2013AA040601).

References

- [1] Seide F, Li G, Yu D. Conversational Speech Transcription Using Context-Dependent Deep Neural Networks. Interspeech. 2011: 437-440.
- [2] Mohamed A, Yu D, Deng L. Investigation of full-sequence training of deep belief networks for speech recognition. Interspeech. 2010: 2846-2849.
- [3] Smith J R, Chang S F. VisualSEEK: a fully automated content-based image query system. Proceedings of the fourth ACM international conference on Multimedia. 1997: 87-98.
- [4] Gotlieb C C, Kreyszig H E. Texture descriptors based on co-occurrence matrices. Computer Vision, Graphics and Image Processing, 1990, 51(1): 70-86.
- [5] Hough P V C. Method and means for recognizing complex patterns: U.S. Patent 3,069,654. 1962-12-18.
- [6] Cross G R, Jain A K. Markov random field texture models. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 1983: 25-39.
- [7] Gibbs J W. The collected works of J. Willard Gibbs[M]. Longmans, Green and Company, 1928.
- [9] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems. 2012: 1097-1105.
- [10] Do C, Ng A Y. Transfer learning for text classification[C]. NIPS. 2005.
- [11] Hinton G, Osindero S, Teh Y W. A fast learning algorithm for deep belief nets[J]. Neural computation, 2006, 18(7): 1527-1554.
- [12] Van de Sande K E A, Uijlings J R R, Gevers T, et al. Segmentation as selective search for object recognition. IEEE International Conference on Computer Vision. IEEE, 2011: 1879-1886.