ISSN 2616-5775 Vol. 4, Issue 7: 46-51, DOI: 10.25236/AJCIS.2021.040707

Automatic Annotation Method of Terror Image Based on Integrated Deep Migration Learning

Wenjun Zhao[1,a], Miaolei Deng[1,b], Dexian Zhang[1,c,*], Hui Gao[2,d]

[1]School of Information Science and Engineering Henan University of Technology, Zhengzhou, China
[2]School of Mechanical and Electrical Engineering Henan University of Technology, Zhengzhou, China
[a]841804007@qq.com, [b]dmlei2003@163.com, [c]zdx@haut.edu.cn, [d]ghshow@139.com
**Dexian Zhang*

Abstract: *With the continuous development of Internet technology, anti-terrorism images have become an important part of network anti-terrorism. In this paper, a method of automatic annotation of terror images based on integrated deep migration learning is proposed by combining parameter migration and ensemble learning, which can help filter terror information in web pages. Firstly, the deep convolution neural network model is used to train the terror image model in the source domain, and then the migration from the source domain to the target domain is realized through the migration learning technology. Then, the integrated learning framework is used to integrate the transfer learning model. Experimental results show that the accuracy and recall rate of the proposed algorithm are obviously improved.*

Keywords: *Terror image, automatic labeling, transfer learning, integrated learning*

1. Introduction

The images of terrorism illegally spread on the Internet have become an important inducement of frequent terrorist cases and one of the biggest sources of poison affecting social stability and security. International terrorist organizations use the Internet to spread terrorist thoughts, develop terrorist attack methods such as organization members, command and plan terrorist activities, and carry out cyber-attacks against Internet infrastructure, which poses a serious threat to the network security of countries all over the world. Therefore, it is extremely urgent to strengthen the supervision of cyber terrorism.

In recent years, with the rapid development of artificial intelligence and image processing technology, image classification and image detection technology based on deep migration learning technology gradually mature, but there are few research results on terror image recognition. Therefore, from the perspective of parameter migration and ensemble learning, this paper proposes an automatic annotation method of terror image based on information fusion and integrated deep migration learning.

2. Related Research

At present, the research on video recognition technology of terrorism in China includes two aspects: video image and voice, among which the research on image occupies the main position. Fu Yabin[1] proposed a fast template matching algorithm for Logo detection of horror video. The template matching speed was improved by coding gray values for rough matching and then using phase correlation for fine matching. Liu Wei[2]uses MPGE-7 audio and visual descriptors for video detection of terrorist violence, and adopts a video detection model based on BP(Back Propagation) network, which obviously improves the video detection effect. Generally speaking, there are few researches on the analysis and recognition of terrorist images.

In recent years, deep learning models have become a research hotspot in the field of computer vision. Deep learning models shine brilliantly in ILSVRC (ImageNet Large Scale Visual Recognition Challenge), and image annotation can be classified into multi-classification problems of images from the perspective of classification. Therefore, the structures of these deep learning models are mostly suitable for image annotation tasks. Wu Baoyuan[3,4] introduced a determinant point process (DPP[7]) based on VGG(Visual geometry group)[5] network and Gan (Generative Adversarial Nets)[6] network, respectively, and retrieved k subsets of marks from all possible marks as the final image annotation.

ISSN 2616-5775 Vol. 4, Issue 7: 46-51, DOI: 10.25236/AJCIS.2021.040707

Zhang Junjie[8] carries out semantic annotation on social networking images, which is introduced into ResNet[9] network to further improve the annotation accuracy of images with inconspicuous visual features. Wang jiang[10] combined RNN(recurrent neural networks) model with CNN (Convective Neural Networks) model to form a unified framework for modeling and extracting the correlation between tags.

Because traditional deep learning models are mostly designed for the classification tasks of large data sets in specific fields, redesign or training the network will cause a series of problems such as over-fitting of the network and increasing the computational burden in the face of new field visual tasks, small data sets or new data sets with fewer tag categories. Therefore, some scholars put forward the idea of deep transfer learning to apply the knowledge or pattern learned in a certain field or task to different fields or problems, such as literature[11-15]. Chen Mengfu[16] designed a deep neural network model and migration learning method to accurately classify and identify Internet images. Yan Liang [17] proposed an automatic labeling method for terrorist images based on ensemble classification, which trained several sub-networks by means of migration learning, and then fused the outputs of sub-networks by ensemble learning, thus improving the problem of large differences in model labeling accuracy among different label categories caused by sample imbalance.

3. Algorithm Design

3.1. CNN Model Design

Convolution neural network is selected in this paper. AlexNet[18] has small number of parameters, fast training speed and good classification accuracy on a small number of image databases. Therefore, this paper applies the parameters trained by AlexNet in ImageNet classification task to terrorist image annotation task. The input of AlexNet model is RGB image, which includes five convolution layers and three fully connected layers. Finally, the prediction is output by softmax function, and the model is trained by the loss function established by prediction results and labels. CNN model adopts sigmoid function to face the gradient dispersion problem when the network is deep, and ReLU function is defined as:

$$f(x) = \max(0, x) \tag{1}$$

The AlexNet model uses GPU for parallel computing to speed up the training process.

3.2. Deep Migration Learning Network Design

Image annotation is a multi-label learning process. First, a neural network classifier needs to be pre-trained on the training set of the source domain, then some parameters are migrated to the target domain by model migration. Finally, the classifier is retrained and upgraded by using the labeled data of the target domain, and the classification task of panic images of the target domain data is completed. The image annotation model based on deep migration learning proposed in this paper is shown in Figure 1.

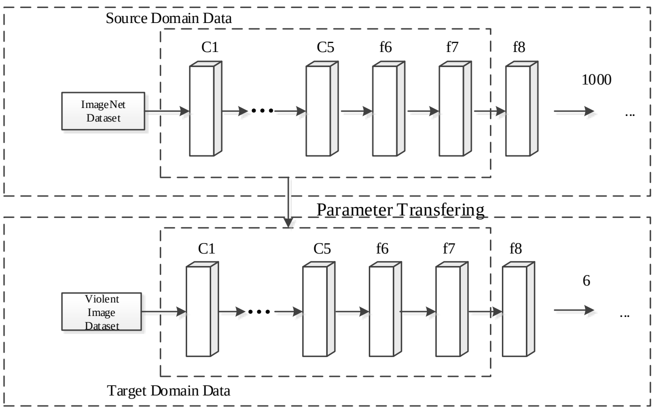

Figure 1: Image annotation model of deep migration learning.

In this model, C1~C5 is a five-layer convolution process, and f6~f8 is a fully connected layer. In the first subtask, the input source task is trained by deep learning layer to obtain image classification,

ISSN 2616-5775 Vol. 4, Issue 7: 46-51, DOI: 10.25236/AJCIS.2021.040707

and C1~C5, f6, f7 migrate the results of the layer to the second subtask. After training, the result of f7 layer is a high-dimensional vector, which corresponds to the features extracted in the middle. After receiving the f7 result, f8 layer performs dimension reduction operation to obtain the multi-label of the second subtask. In the second subtask, the image annotation results are obtained by training the migrated parameters.

3.3. Integrated Deep Learning Framework

In order to make full use of information from multiple sources, this paper proposes a new integrated deep migration learning framework. As shown in Figure 2, this framework mainly consists of two parts: Multi-domain Joint Transfer Learning and Classification Stage.

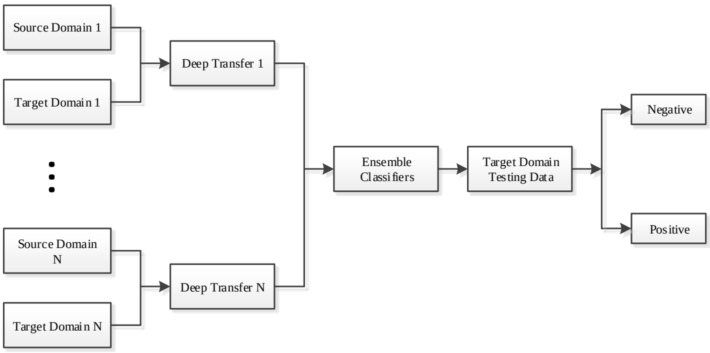

Figure 2: Framework of deep Ensemble Learning model.

There are n source domains {D1, D2, …, Dn}. firstly, the source domain DiT is combined with the target domain labeled data set DT, and deep migration learning is performed on each combined block [DiT, DT] to obtain the target domain In the testing stage of the target domain, the test samples of the target domain are input, the labels of the test samples are predicted by the trained weighted ensemble classifier, and the terror picture labels are output.

4. Experimental Results and Analysis

4.1. Data Set and Experimental Environment

(a) Violent examples

(b) Non-violent examples

Figure 3: Sample of Terrorist Image Dataset.

According to the current research, there is no data set for violent terrorist video in the public domain, and the training and testing of the experimental model in this paper needs a large number of violent terrorist images. Therefore, in this paper, crawler technology is used to crawl the relevant images from specific web pages, and then the data is supplemented by manual search. The common data set and self-made data set are used for training and testing. First, the public dataset ImageNet vid

800 is used, with 500 images in each class. Secondly, in order to obtain more violent terrorist data, this paper uses the method of data enhancement to clip, classify and organize the collected video into 200 subsets, each of which has 300 pictures. Finally, the 1000 subsets were divided into six sub categories, namely, specific dress wearer, gun holder, specific logo, bearded, black veiled and normal human. A total of 460000 images were divided into training set and data set according to the ratio of 4:1.Some examples of violent and non-violent terrorist samples in the violent terrorist image dataset are shown in Figure 3 (a) and Figure 3 (b), respectively.

The algorithm in this paper is implemented by Tensorflow 1.4.0 framework and python programming under the Ubuntu 16.04 operating system. The algorithm is trained and tested on Intel Broadwell 2.4 GHz processor, and accelerated by GPU (Tesla V100).

4.2. Evaluation Index

In this paper, according to the idea of classification, we choose the common evaluation indexes of image annotation algorithm, such as precision P, recall rate R, F1, N +, average precision (AP), mean average precision (map) to test the advantages and disadvantages of various models. The calculation formula is as follows:

$$\mathrm{P} = \frac{1}{M}\sum_{k=1}^{M}\frac{Correct(\omega^k)}{Predicted(\omega^k)} \tag{2}$$

$$\mathrm{R} = \frac{1}{M}\sum_{k=1}^{M}\frac{Correct(\omega^k)}{GroundTruth(\omega^k)} \tag{3}$$

$$\mathrm{F1} = \frac{2PR}{P+R} \tag{4}$$

Among them, $Correct(\omega^k)$Indicates the correct number of images to predict for the kth tag, $Predicted(\omega^k)$ Represents the total number of predicted images for the kth tag, $GroundTruth(\omega^k)$Represents the total number of images labeled with the kth label.F1 combines the results of P and R. when the value of F1 is higher, the model method is more ideal.

4.3. Experimental Results and Analysis

In order to show the effect of transfer learning, the accuracy of the model without transfer learning is added as a comparison. As shown in Figure 4, the horizontal axis is the training iteration period, and the vertical axis is the average classification accuracy on the validation data set.

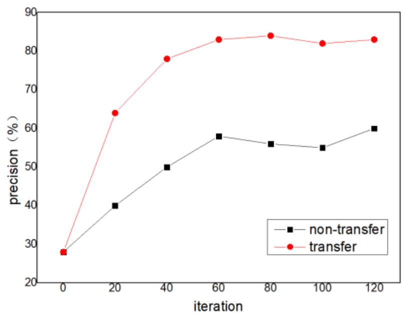

Figure 4: Results of Model Training.

Table 1: Performance comparison of image annotation methods.

Model	P/%	R%	F1%
MBRM	24	25	24
HMM	19	18	18
LSA	7	9	8
KNN+CCA[22]	42	52	46
KSVMMN[23]	82.2	61.1	70.8
CNN+RNN[24]	80	72	75.7
Ours	83	70	76

ISSN 2616-5775 Vol. 4, Issue 7: 46-51, DOI: 10.25236/AJCIS.2021.040707

In order to verify the effectiveness of the proposed algorithm, this paper selects the correlation model MBRM[19], hidden Markov model HMM[20], topic model LSA[21], nearest neighbor model KNN, SVM model ksvmmn, CCA, CNN and RNN algorithm. In this paper, the single algorithm is compared with the integrated algorithm, and the results show that compared with the single network, the integrated learning has a great improvement in the accuracy of terrorist image annotation method. Table 1 shows the accuracy and recall rate of several algorithms in the violent terrorist image data set. Compared with the classic ksvmmn algorithm, the accuracy rate is improved by 0.8%, and the recall rate is increased by 8.9%.

5. Conclusions

In this paper, we propose an automatic labeling method for violent terrorist images based on integrated deep transfer learning. By using convolutional neural network, transfer learning and ensemble learning strategies, the model transfer from source domain to target domain is realized by transfer learning. Finally, the output of classifier is combined with ensemble learning method, it improves the problem that uneven samples will lead to the big difference of labeling accuracy in each label category. In addition, the number and diversity of training data need to be further expanded, and network model and training method are also the problems to be studied in the next step.

Acknowledgments

This work was supported by National Key R&D Program of China2018******02.

Zhao Wenjun, born in 1993, is a doctoral candidate and mainly researches intelligent information technology. E-mail: 841804007@qq.com.

Deng Miaolei, born in 1977, holds a PhD. degree and is an associate professor, mainly engaged in intelligent information technology.

Corresponding author: Professor Zhang Dexian, born in 1961, is a doctoral supervisor with a PhD. degree. He is also the vice-chairman of Information and Automation Sub-association of Chinese Cereals and Oils Association (CCOA), a scientific and technological innovation talent in Henan Province and an academic and technical leader of the Education Department of Henan Province. He mainly studies intelligent information technology, data mining and machine learning.

Gao Hui, born in 1983, is a doctoral candidate and mainly researches intelligent information technology.

References

[1] FU Yabin. Design and implementation of the Violent-terrorist video recognition system based on Logo markers detection[D].Beijing: Beijing Jiaotong University,2016:15-30.
[2] Liu Wei. Video detection of terrorist violence based on MPEG-7 audio-visual features [D]. Hebei University of technology, 2014
[3] Wu Baoyuan, Jia Fan, Liu Wei, et al. Diverse image annotation[C] //Proc of the IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ ; IEEE, 2017:2559-2567.
[4] Wu Baoyuan, Chen Weidong, Sun Peng, et al. Tagging like humans: Diverse and distinct image annotation[C] //Proc of the IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ ; IEEE, 2018:7967-7975.
[5] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition [J]. arXiv preprint, arXiv:1409.1556,2014.
[6] Goodfellow I, Pouget-abadie J, Mirza M, et al. Generative adversarial nets [C]//Proc of Advances in Neural Information Processing Systems. Cambridge, MA; MIT, 2014: 2672-2680.
[7] Kulesza A, Taskar B. Determinatal point processes for machine learning[J]. Foundations and Trends in Machine Learning, 2012,5(2/3): 123-286.
[8] Zhang Junjie, Wu Qi, Zhang Jian, et al. Mind your neighbours: Image annotation with metadata neighbourhood graph coattention networks[C] //Proc of IEEE Conf on Computer Vision and Pattern Recognition .Piscataway, NJ ; IEEE, 2019:2956-2964.
[9] He Kaiming, Zhang Xiangyu, Ren Shaoqing, et al. Deep residual learning for image recognition[C] //Procof IEEE Conf on Computer Vision and Pattern Recognition (CVPR).Piscataway, NJ ; IEEE,

ISSN 2616-5775 Vol. 4, Issue 7: 46-51, DOI: 10.25236/AJCIS.2021.040707

2016:770-778.

[10] Wang Jiang, Yang Yi, Mao Junhua, et al. CNN-RNN:A unified framework for multi-label image classification[C] //Proc of IEEE Conf on Computer Vision and Pattern Recognition .Piscataway, NJ ; IEEE, 2016:2285-2294.

[11] Oquab M, Bottou L, Laptev I, et al. Learning and transferring mid-level image representations using convolutional neural networks[C] //Proc of IEEE Conf on Computer Vision and Pattern Recognition .Piscataway, NJ ; IEEE, 2014:1717-1724.

[12] Yosinski J,,Clune J,Bengio Y, et al. How transferable are features in deep neural networks [C] //Proc of Annual Conf on Neural Information Processing Systems. Cambridge, MA; MIT, 2014: 3320-3328.

[13] Gong Yunchao, Jia Yangqing, Leung T, et al. Deep convolutional ranking for multilabel image annotation[J]. arXiv preprint, arXiv:1312.4897,2013.

[14] Tan Ben, Song Yangqiu, Zhong Erheng, et al. Transitive transfer learning[C] //Proc of the 21st ACM SIGKDD Int Conf on Knowledge Discoverry and Data Mining. New York: ACM,2015: 1155-1164

[15] Tan Ben, Zhang Yu, Pan S, et al. Distant domain transfer learning [C]//Proc of the 31st AAAI Conf on Artifitial Intelligence. Menlo Park, CA: AAAI, 2017:2604-2610.

[16] Chen Mengfu. Research on automatic recognition model for terrorism related image based on Transfer Learning,2020,46(09):1677-1681.

[17] Yan Liang. Violet image annotation using ensemble learning [J]. Terahertz Journal of science and electronic information, 2020,18 (02): 306-312

[18] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[J]. Advances in neural information processing systems, 2012, 25: 1097-1105.

[19] Feng S L, Manmatha R, Lavrenko V. Multiple bernoulli relevance models for image and video annotation[C]//Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004. IEEE, 2004, 2: II-II.

[20] Ghoshal A, Ircing P, Khudanpur S. Hidden Markov models for automatic annotation and content-based retrieval of images and video[C]//Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval. 2005: 544-551.

[21] Pham T T, Maillot N E, Lim J H, et al. Latent semantic fusion model for image retrieval and annotation[C]//Proceedings of the sixteenth ACM conference on Conference on information and knowledge management. 2007: 439-444.

[22] Murthy V N, Maji S, Manmatha R. Automatic image annotation using deep learning representations[C]//Proceedings of the 5th ACM on International Conference on Multimedia Retrieval. 2015: 603-606.

[23] Liu Y, Wen K, Gao Q, et al. SVM based multi-label learning with missing labels for image annotation[J]. Pattern Recognition, 2018, 78: 307-317.

[24] 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.