An Accident Analysis and Prevention Model based on Heterogeneous Convolution Kernel

ZhipengYang¹, ShashaWang¹ and YanlingLi¹⁺

¹ School of Computer and Information Technology, Xinyang Normal University, Xinyang 464000, China

Abstract. In order to reduce the incidence of safety accidents in factories, an accident analysis and prevention model based on heterogeneous convolution kernel is proposed. On the basis of the framework of convolutional neural network, the traditional fixed convolutional kernel optimization network was replaced by constructing heterogeneous convolutional kernel, which was applied to the industrial workshop to monitor the age and gender of workers in the workshop in real time, so as to reduce the calculation amount and the number of parameters while maintaining the efficiency of characterization. Through the comparative experiment in the standard database of Adience and SCUT-FBP 5500, the experimental results of various methods are analyzed. The results show that the model proposed in this paper can achieve 92% accuracy in gender classification of characters and 67% accuracy in age prediction. It can monitor the gender and age of workers in the workshop, real-time warn the potential safety hazards, and reduce the safety of the factory to a certain extent the occurrence rate of all accidents.

Keywords: Heterogeneous Convolution Kernel; Convolutional Neural Network; Age Predicts; Gender Prediction; Accident analysis; Accident prevention.

1. Introduction

In factories, the production process is complex, and the level of mechanization is low by using mechanical and manual collaborative production mode. Production process involves a variety of hazardous production equipment and facilities, lack of safety protection equipment and facilities, and part of the production process because of its industry characteristics, the dangerous and harmful factors involved have a certain particularity, under certain conditions easy to trigger into accidents, thereby endangering the lives of employees and causing property losses of enterprises. With the rapid development of the era of big data, people continue to try and innovate on the research of machine with the same ability of human vision. How to make better use of the machine to analyse the characteristics of workshop workers in the factory, so as to get the distribution of workshop population, and provide a reliable basis for the intelligent management of the factory. According to the crowd distribution in the workshop, the on-the –job situation of workers can be monitored in real time. More importantly, the location of workers can be effectively restricted. If workers enter areas that are not allowed to enter in this age group, they will give alarm warnings, which can effectively reduce the occurrence of factory safety accidents to a certain extent.

In this paper, a convolution neural network model based on heterogeneous convolution kernel is constructed, which aims to classify gender and age accurately according to facial features, so as to statistics the distribution of workshop population, and urge workshop population, and urge workshop workers not to skip jobs, not to go to areas where their age or gender should not enter. It has important significance in theory and practical application.

2. Related Work

2.1. Convolutional Neural Network

Corresponding author. Tel.: + 0376-6393167.
E-mail address: 1y175@163.com.

Convolutional neural networks (CNN) have achieved excellent performance in both visual and natural language processing, and the overall trend to further improve performance has made the model more complex and deeper. However, using deeper networks to improve accuracy by increasing model complexity is costly, and computing costs (FLOPS) will increase dramatically. Therefore, in order to reduce FLOPS and make the model more efficient, researchers have proposed various types of convolution filters.

Most recent architectures have used a combination of these convolution filters to produce efficient models. Many common models also use these convolutions to explore new architectures that can reduce FLOPS. Another common method to improve model efficiency is compression model, which can be roughly divided into three categories: connection pruning, filter pruning [1-3] and quantization [4]. There are two different goals for using efficient convolution filters. The focus of one kind of research is to design the smallest FLOPS architecture and compromise on accuracy. Another kind of research focuses on improving the accuracy while ensuring that FLOPS of the model are the same as the original architecture. Different from these two methods, our approach focuses on reducing FLOPS for a given model or architecture by using a new convolution kernel without losing accuracy. Experiments show that our proposed method has lower FLOPS and maintains the accuracy of the underlying model or architecture. In our proposed model, we choose a different strategy to improve the efficiency of the existing model without sacrificing its accuracy. Architecture search methods take years of research to get an optimal architecture. Therefore, instead of using a new efficient architecture, we use an efficient convolution operation (convolution filter), namely heterogeneous convolution.

2.2. Relative Relation

In order to better understand the relationship between age, gender and the occurrence of safety accidents in a factory [5], a statistical analysis of safety accidents in a factory in HuNan Province was carried out. Statistics show that the vast majority of accidents are closely related to employees' weak sense of safety. They go in and out of places where they are not supposed to be in and out of their age and do not stick to their posts. Statistical data show that the proportion of female and male deaths is 66.7% and 33.3% respectively. The proportion of males was 92.7% and 87.1% respectively in the serious and minor injuries, accounting for the vast majority, the specific description is may be depicted as in Figure 1.



Fig. 1: This is the distribution of sex and age structure of casualties.

In view of this situation, this paper designs a convolution neural network model to accurately classify the gender and age of workshop population, and monitors the position of workshop workers in real time. Once it is found that someone enters the area that is not allowed to enter in this age group, the alarm will be activated and a warning will be issued. As far as the current design concept is concerned, in theory, it can effectively protect the personal safety of workshop workers and reduce the incidence of safety accidents. It also of great value in practical application.

3. Methodology

3.1. Design of Proposed Model

The specific problems to be solved by the proposed model are similar to those of classification. Among them, there are 2 in gender classification (male and female), and 8 in age classification (using 8 age segments).

The input image is represented by I_t , $t = 1, 2, 3, \dots, T$, and the score vector is $x_t = W\phi(I_t) + b$, where $W \in \mathbb{R}^{(N \times D)}$, $b \in \mathbb{R}^N$, the number of neurons in the full connection layer in the last layer of the network is N, and the data label is represented by $c_t \in \{1, 2, 3, \dots, N\}$. The SoftMax loss function is calculated according to the score vector and data label

$$E(\phi) = -\sum_{t} \log\left(e^{\langle e_{q_{t}}, x_{t} \rangle} / \sum_{q=1,2,3,\cdots,N} e^{\langle e_{q}, x_{t} \rangle}\right)$$
(1)

Where $x_t = \phi(I_t)$, the sparse representation of classification *c* is e_t .

3.1.1. Gender Classification Model

Because gender prediction based on face image belongs to the two-classification problem in the field of pattern recognition and classification, many researchers use VGG [6] network to classify the features extracted from face regions. However, because of the complexity of VGG network model, the resource occupancy rate is too high and the flexibility of the model is not very high. Therefore, the proposed model replaces the fixed-size convolution cores in the network and uses heterogeneous convolution cores to optimize the network, which can guarantee the same accuracy of the original model and reduce FLOPS. Figure 2 shows the characteristics of the heterogeneous convolution filters. Among them, M refers to the input depth (the number of input channels), and P is a set parameter (controlling the number of different types of cores in the convolution filter). In M nuclei, the size of M/P nuclei is 3×3 , and the rest are 1×1 .



Fig. 2: Standard convolution filters (Isomorphic) and Heterogeneous convolution filters (HetConv).

The convolutional neural network model used in this paper is shown in Figure 3. There are four convolution layers and four pooling layers in the network. The activation function of the final full connection layer is ReLu. The parameters of the network structure are set as follows:

- The input of the network is RGB image of any size.
- Each convolution layer uses a convolution core with size of 3×3 and 1×1 , parameter P is 4, step size is 1, sampling window of 2×2 is used in the pooling layer, maximum mean pooling method is adopted, step size is fixed at 2.
- The first convolution layer and the pool layer have the same number of cores as 32, the second convolution layer and the pool layer have the same number of cores as 64, the third convolution layer and the pool layer have the same number of parameters as 128, the last convolution layer and the pool layer have the same number of cores as 48, the parameters of the full connection layer are set to 1024, and the output classification value is 2.
- Because different padding and step size will affect the final image output size, this paper uses the same way to complete with all 0, so as to ensure that the length and width of the final convolution results remain the same as the input.

3.1.2. Age Classification Model

The subtle differences of images exOtracted from different age groups will result in different results because of the angle of image data acquisition or the wearing of jewelry. In addition, the workshop lighting is different from the standard laboratory environment, which requires higher classification model for human age groups than for gender classification. Classical age classification network models generally adopt VGG

network, and some relevant scholars [7] preprocessed human faces by combining VGG with SVM to improve the accuracy of age prediction, but greatly reduced the computational efficiency. In this paper, in order to meet the flexibility and efficiency of the application in a complex environment, a VGG-Face age classification model that has been balanced between accuracy and calculation amount was selected, and a heterogeneous convolution kernel was used in the convolution kernel.



Fig. 3: Gender classification model of convolutional neural network based on heterogeneous convolution kernel.

The age classification model used in this paper is shown in Figure 4. The first 16 layers of VGG network structure are adopted in the network model. However, in order to make the input pictures better simulate the processing of human eyes, we need to go through several levels of convolution operation. The simulation cognitive process will make the network parameters too much, and the training speed will greatly slow down. The model uses pre-training instead of the traditional gradient descent method to accelerate the training speed of the network. The parameters of the network structure are set as follows:

- The pooling layer in VGG network divides the network convolution layer into five layers.
- The convolution kernel used in network convolution operation is similar to the gender classification model. The convolution kernel size is 3×3 and 1×1. The parameter P is 4 and the step size is 1. The maximum mean pooling method is used in the pooling layer. The size of the sampling window is 2×2 and the step size is fixed at 2.
- The number of convolution cores in all layers is the same. The number of convolution and pooling cores in the first layer is 64, the number of convolution and pooling cores in the second layer is 128, the number of convolution cores in the third layer is 256, the number of convolution cores in the fourth layer is 512, the number of full connection cores in the last layer is 2000, and the final output forecast value is 8 categories.



3.1.3. Accident Analysis and Prevention Model

On the basis of convolution neural network, an accident analysis and prevention model based on heterogeneous convolution kernels is proposed, which integrates age and gender prediction to monitor the crowd distribution in workshop in real time. This model is a combination of gender classification model and age classification model. The purpose of this model is to well locate the distribution of the workshop population, analyze the safety accidents in the workshop for many years, learn experience, and strictly limit the position of workers of different ages and genders. Once they enter the inadmissible areas, the model will give an alarm. This is designed to reduce the occurrence of safety accidents. The Figure 5 gives the specific work-flow diagram of the model. In the figure, the operation of the whole model is introduced in detail, including real-time detection of factory population, uploading image data, prediction of age and gender, comparative analysis with previous cases, warning or prompting workers, which can reduce the occurrence of safety accidents in the workshop to some extent.



Fig. 5: The specific work-flow diagram of the model.

4. Experiment and Results

The experiment is implemented by TensorFlow on Lenovo desktop computer with i7-6700CPU, 8G memory and windows operating system. The proposed model is validated on two data sets, and the accuracy of prediction is recorded on different data sets. The comparison experiment with other models is also done.

Sex prediction experiments were carried out on Adience data set [7] and SCUT-FBP5500 data set respectively. Before image input network training, preprocessing was needed, and the image was scaled to 256×256 pixels. The experiment compares the model with CNN+SVM [8], Ft-VGG-Face [9], Ft-VGG-Face+SVM [9] on two datasets. The experimental results are as follows.

Table 1: Comparison of Correct Prediction Rate on Adience Dataset		Table 2: Comparison of Correct Prediction Rate onSCUT-FBP5500 Dataset	
Method	Precision	Method	Precision
CNN+SVM	87.2%	CNN+SVM	85.2%
Ft-VGG-Face	92.0%	Ft-VGG-Face	86.2%
Ft-VGG-Face+SVM	91.9%	Ft-VGG-Face+SVM	89.9%
The Model Proposed in this Pape	93.8%	The Model Proposed in this Pape	90.6%

According to table 1 and table 2, the experimental accuracy of the proposed model on Adience Dataset was 93.8%, while that on SCUT-FBP5500 Dataset was 90.6%. Compared with the former, the latter has a slightly lower accuracy. However, in general, the accuracy of the model proposed in this paper is slightly higher than that of other models, whether it is conducted on Adience Dataset or SCUT-FBP5500 Dataset. Therefore, the model proposed in this paper that uses heterogeneous convolution kernel for gender classification is no less accurate than other models, with an average accuracy of 92.2%.

When the task of age classification is validated on Adience data set and SCUT-FBP5500 data set, the face data set is pre-processed into 256×256 pixels. The model is compared with VGG16-DEX-IMDB [10], CNN [11], VGG-Face+SVM [9]. The experimental results are as follows.

Table 3: Comparison of Correct Prediction Rate on Adience Dataset		Table 4: Comparison of Correct Prediction Rate on SCUT-FBP5500 Dataset	
Method	Precision	Method	Precision
VGG16-DEX-IMDB	65.2%	VGG16-DEX-IMDB	62.2%
CNN	56.2%	CNN	59.2%
VGG-Face+SVM	57.9%	VGG-Face+SVM	56.9%
The Model Proposed in this Paper	66.4%	The Model Proposed in this Paper	68.4%

Table 3 and 4 show that the experimental accuracy of the model for age classification is 66.4% on Adience Dataset and 68.4% on SCUT-FBP5500 Dataset. Compared with the former, the latter has a slightly higher accuracy, but in general, the accuracy of the model proposed in this paper is slightly higher than that of other models, whether it is experimented on Adience Dataset or SCUT-FBP5500 Dataset. Therefore, the accuracy of the model proposed in this paper is no less than that of other models, with an average accuracy rate of 67.4%.

5. Conclusions

The proposed model is based on the business background of personal safety analysis in the factory and accurate positioning of population distribution, and integrates age and gender prediction to put forward an accident analysis and prevention model. The model uses heterogeneous convolution kernel in the convolution layer, which greatly reduces the flop and ensures the accuracy. This model can monitor the workshop population in real time, accurately give the distribution of people of different ages and genders in the workshop, analyze safety accidents, strictly divide the safe working area of age groups and genders, prevent the recurrence of safety accidents, and reduce the incidence of safety accidents in the factory to some extent. At the same time, the experimental results show that the method proposed in this paper is superior to other methods mentioned in this paper, which is theoretically enough to provide help for preventing safety problems in factories. There are still many improvements in the model, simplify the network structure, and improve the accuracy of the model to apply to other broader fields.

6. Acknowledgments

We acknowledge Professor YanlingLi for his supervision and guidance in this research.

7. References

- [1] Singh, P.; Manikandan, R.; Matiyali, N.; Namboodiri, V. Multi-layer Pruning Framework for Compressing Single Shot MultiBox Detector. CoRR. 2018, abs, 1811.08342.
- [2] Singh, P.; Kadi, V.; Verma, N.; Namboodiri, V. Stability Based Filter Pruning for Accelerating Deep CNNs. CoRR. 2018, abs, 1811.08321.
- [3] Singh, P.; Verma, V.; Rai, P.; Namboodiri, V. Leveraging Filter Correlations for Deep Model Compression. CoRR. 2018, abs, 1811.10559.
- [4] Rastegari, M.; Ordonez, V.; Redmon, J.; Farhadi, A. XNOR-Net: ImageNet Classification Using Binary Convolutional Neural Networks. CoRR. 2016, abs, 1603.05279.
- [5] Deng, S. Statistical analysis and safety evaluation of production accident in casting workshop of a machinery manufacturing enterprise in HuNan province. 2017.
- [6] Kuang, Z.; Huang, C.; Zhang, W. Deeply Learned Rich Coding for Cross-Dataset Facial Age Estimation. In 2015 IEEE International Conference on Computer Vision Workshop, ICCV Workshops 2015, Santiago, Chile, December 7-13, 2015; IEEE Computer Society; Publishing House: Santiago, Chile, 2015; pp. 338–343.
- [7] Eidinger, E.; Enbar, R.; Hassner, T. Age and gender estimation of unfiltered faces. IEEE Trans. Information Forensics and Security. 2014, 12, 2170–2179.
- [8] Wolfshaar, J.; Karaaba, M.; Wiering, M.; Deep Convolutional Neural Networks and Support Vector Machines for Gender Recognition. In IEEE Symposium Series on Computational Intelligence, SSCI 2015, Cape Town, South Africa, December 7-10, 2015; IEEE; Publishing House: Cape Town, South Africa, 2015; pp. 188–195.
- [9] Ozbulak, G.; Aytar, Y.; Ekenel, H. How Transferable are CNN-based Features for Age and Gender Classification? CoRR. 2016, abs, 1610.00134.
- [10] Rothe, R.; Timofte, R.; Van, L. Deep Expectation of Real and Apparent Age from a Single Image Without Facial Landmarks. International Journal of Computer Vision. 2018, 2–4, 144–157.
- [11] Ranjan, R.; Zhou, S.; Chen, J.; Kumar, A.; Alavi, A.; Patel, V.; Chellappa, R. Unconstrained Age Estimation with Deep Convolutional Neural Networks. 2015 IEEE International Conference on Computer Vision Workshop, ICCV Workshops 2015, Santiago, Chile, December 7-13, 2015; IEEE Computer Society; Publishing House: Santiago, Chile, 2015; pp. 351–359.