Dangerous goods identification based on multi-channel neural network

XiaoTian Wei¹*, ZiQiang Hao²*, Bo Du³

¹ Department of Control Engineering, Changchun University of Science and Technology, Changchun, China.
²*,³752258410@qq.com, 7663410@qq.com, 742852677@qq.com

Abstract- In the current society, there is an increasing demand for dangerous goods identification technology in X-ray images, but at the current stage, most of the identification of dangerous goods in X-ray images still relies on artificial eye recognition. In order to solve this problem, this paper proposes a method for automatically and intelligently identifying dangerous goods in X-ray images based on the transformation of the convolutional neural network. By adding multi-channel convolution and normalization to the convolutional neural network, the target features are extracted to achieve automatic detection of dangerous goods. The purpose of better identification. In the simulation experiment, using the public data set and self-built data set in the X-ray security inspection field, the accuracy of the identification of dangerous goods in the X-ray image was obtained more satisfactory results than the traditional dangerous goods identification. The improved Alex Net network’s testing accuracy on contraband knives and guns is 8.53% and 11.6% higher than the training accuracy of the original Alex Net network.

1. Introduction

In today’s increasingly severe security situation, in order to help staff accurately identify the types of items in the luggage, so as to confirm whether passengers carry dangerous goods such as knives and guns in their hand luggage, reduce the probability of dangerous incidents, and ensure the safety of passengers, it is necessary to improve X Light image recognition effect. At present, the recognition of X-ray images is mainly through the recognition of workers’ eyes. With the development of artificial intelligence, the use of artificial intelligence algorithms to automatically detect and recognize special objects from X-ray images is a future development trend. The processing method of X-ray security images is not exactly the same as that of visible light images, which is caused by its own imaging and collection methods. It mainly involves image smoothing, image enhancement, image denoising, etc. According to its own imaging characteristics, it is necessary to conduct in-depth research on processing methods such as the extraction of independent objects in the image, the separation of overlapping objects, and the identification of object materials. According to the above, an automatic detection system for prohibited items can be established based on the characteristics of prohibited items to further improve the efficiency of dangerous goods identification.

With the introduction of some new theories and methods and the improvement of application technologies, such as wavelet theory, statistical theory, neural network and other methods are gradually applied to the processing of X-ray images, Some dangerous goods identification technologies are slowly developed. However, there is still a lot of research space for the X-ray image processing technology of neural network. However, in the past ten years, many domestic upgraded and
imported equipments have average functional quality for X-ray image processing and recognition. Therefore, the intelligent image diagnosis system with high reliability and high recognition rate needs further research\cite{1}\cite{2}. Begin to pay more attention to the processing and research of X-ray images.

In 2006, in response to the problem of object occlusion in X-ray images, Ge Jipeng et al. proposed a method to remove the overlap effect in X-ray images by extracting the true gray levels of objects. In 2006, Yuan Peixin established a material classification and recognition curve under dual-energy X-ray fluoroscopy conditions, and proposed an object recognition method based on edge operators for the overlap of objects in X images. In 2009, He Xiuping et al. combined fuzzy theory to realize an X-ray image enhancement method. In 2013, Wang Huiying et al. studied the characteristics of X-ray backscatter images and proposed a method for segmentation of contraband images based on EM clustering\cite{3}. In 2015, S. Ren et al. proposed the Faster RCNN detector and introduced a regional recommendation network, which can achieve regional recommendation with almost no additional consumption, but the detection speed is still not ideal. In 2019, Liu Kun, Wang Dian, and Rong Mengxue jointly proposed an X-ray image classification method based on a semi-supervised semi-generative adversarial network, which improves the learning performance of using labeled data compared with other semi-supervised learning.

In recent years, as a branch of artificial intelligence, deep learning has been widely used in various fields that require target detection due to its faster detection speed and higher detection accuracy in the field of image recognition\cite{4}\cite{5}. The mark recognition of forbidden products in X-ray images used to be done manually. Under the premise of ensuring accuracy, how to improve the recognition speed of forbidden products in X-ray images has become an important issue. The regional convolutional neural network is one of the mainstream frameworks in the current target detection field.

This article uses AlexNet as the basic feature extraction network for forbidden band product recognition model training, but because the size of the convolution kernel of the convolution layer in the AlexNet network is a single size, the generated feature map features are not diverse, then it will cause Insufficient feature extraction, which in turn affects the recognition effect. Therefore, this article introduces multi-channel convolution and normalization to the AlexNet classic network to transform the network, and uses the transformed AlexNet network to extract the features of the dangerous goods in the X-ray image.

2. Related theories

2.1. X-ray image

X-ray is also called roentgen ray. It is an electromagnetic waves with the common properties of electromagnetic waves. X-ray is an invisible electromagnetic waves propagating in a straight line in a uniform and isotropic medium. It has the ability to penetrate objects, but the ability to penetrate different objects is different, and it has the effect of destroying cells\cite{6}. Therefore, X-rays are first used in medical diagnosis as soon as they are discovered.

The formation of X-ray images should meet the following three basic conditions: firstly, X-rays have a certain penetrating power, so that they can penetrate the irradiated tissue structure; secondly, there must be a difference in density and thickness, so that during the penetrating process The amount of X-ray remaining after being absorbed will be different; in the end, the remaining X-ray with difference is still invisible, and the X-ray image with contrast level difference can be obtained through the process of imaging.

This X-ray image has the following characteristics:

(1) It is a black and white image, and the image is composed of images with different gray levels from black to white. Commonly used different densities correspond to black and white: high density corresponds to white, medium density corresponds to gray, and low density corresponds to black.

(2) X-ray image is the total projection of all tissues on the X-ray path, which transforms the three-dimensional into a two-dimensional image. Therefore, the X-ray image is compared with the structure of the human body, resulting in morphological distortion, magnification and overlapping of...
the composite image, so X-ray images are superimposed images.

Traditional X-ray imaging technology uses analog technology. Once the X-ray image is produced, its image quality cannot be further improved, and its information is analog, which is inconvenient for image storage, management, and transmission, which limits its development. By 1997, direct digital X-ray imaging technology appeared. The digitization of X-ray images can not only use various image processing technologies to process images and improve image quality, but also display the obtained images simultaneously through various diagnostic technologies. With the development of science and technology, X-ray images are gradually applied to medical detection, object surveying, and the identification of dangerous objects in security checks. However, at the current stage, most of the recognition of X-ray image objects depends on artificial naked eye recognition and judgment, so we need to find a way to automatically recognize objects in X-ray images.

2.2. AlexNet network

The AlexNet network is a network model proposed by Krizhevsky et al. in the ISLVRC Challenge in 2012. The model has attracted worldwide attention since its inception. The emergence of this network has broken the indolence since Hinton proposed deep learning in 2006. The status quo that is not popular and not of people's attention is a turning point for deep learning to be widely studied and applied by researchers. With the development of deep learning, the AlexNet network is also one of the network models preferred by many researchers. Various improved optimization methods have made the robustness of AlexNet applications in various fields continuously improved. The AlexNet network contains a total of 650,000 neurons and 60 million parameters. The network structure diagram is shown in Figure 2.1: From left to right, it contains a total of eleven layers, including five convolutional layers and two pooling layers, two fully connected layers and a softmax output layer, which are also mixed with a local normalization layer and a Dropout layer.

![Figure 1. AlexNet structure diagram.](image)

The AlexNet network attracted much attention at the time because it has the following advantages:

1. Use Relu activation function. The Relu function does not have soft saturation, and as the network deepens, it will not cause the problem of gradient disappearance; secondly, the Relu function does not involve complex exponential calculations, the amount of calculation is small, and the convergence speed is fast; and the Relu function is more in line with biological neurons Incentive mechanism.

2. AlexNet introduces Dropout in the fully connected layer. Dropout is now one of the most common methods to solve training overfitting, by randomly ignoring some neurons during the training process to enhance the fanhuaxing feature of the model, thereby solving the problem of model overfitting.

3. The design proposes a local normalization layer (LRN). The LRN layer creates a competition mechanism for the activity of local neurons, which makes the value with larger response become relatively larger, and inhibits other neurons with smaller feedback, which enhances the generalization ability of the model.

4. The use of multi-GPU co-training technology accelerates the training speed of the network.
3. Improved AlexNet network

3.1. Defects of the original AlexNet network

Although the AlexNet network has greatly improved compared with the Le Net5 network, and the target recognition effect has also been greatly improved, the classic AlexNet network still has certain shortcomings. The AlexNet structure is relatively simple, and the network model is easier to train, but the structure is simple. It brings certain limitations to its ability to extract features, mainly because the convolution kernel used in the model is relatively single, which is only a pure convolution kernel pooling operation, and the transmission of information can only pass through one branch, the feature extraction ability Relatively weak. The parameter list of the Alex Net network, the last three layers of fully connected parameters are 6*6*256*4096, 4096*4096, 4096*1000, especially the results of the first two have reached tens of millions of levels of data, so huge The number of parameters will inevitably cause the training speed to slow down, and at the same time, it will also lead to the existence of overfitting. Despite the existence of dropout, there is no uniform standard for setting the value of dropout[9]. You must rely on experience or try CNN many times. Training mainly relies on the gradient descent method, using the gradient descent method to help find the smallest loss value, so as to infer and update the weight and bias value. If there is a small change in the first few layers of the network, then when it arrives at the back, this small change will be accumulated and amplified, and the data distribution will change, which will make our loss value become very large, and carry out the weighting and biasing. The update of the value also has a great impact, making the network unable to complete the calculation, and the phenomenon of gradient explosion occurs. This article improves some of the above-mentioned defects.

3.2. Multi-channel convolution

The process of convolution is to convolve the image of one channel, and the original convolution layer uses a single size convolution kernel to convolve the input data to obtain several feature maps, and each layer outputs the number of feature maps. It is the number of convolution kernels in this layer.

The input feature map of the convolutional layer is a multi-channel signal, which corresponds to a multi-channel convolution kernel. One-dimensional convolution operation is performed on each channel separately, in order to connect the independent signals together to achieve the purpose of signal fusion. The multi-channel convolution technology can be regarded as an extension of the original convolution operation, adding several convolution kernels of different sizes to a single convolution layer, and finally connecting them together, which will make the generated feature The map features are more diverse.

Apply the multi-channel convolution technology to the AlexNet network and replace the original third convolution layer. The steps are as follows:

1) Change the depth of the convolution kernel of the first layer from 96 to 64. The depth modification of the convolution kernel is to better combine with the depth of step 2. At the same time, it also has the effect of reducing the number of parameters of the first layer, as shown in Figure 3.1;

![Figure 2](image1.png)

Figure2. Network changes in depth.

2) Use four convolution kernels of different sizes to replace the single size 5*5 convolution kernel of the third layer. The sizes of the four convolution kernels are 1*1, 3*3, 5*5, and 7 respectively. *7;

3) Set the depth of the four convolution kernels to 64 respectively. Since the depth of the original single size 5*5 convolution kernel is 256, we replace it with four convolution kernels of different sizes, so the total depth of the replaced convolution kernel must still be consistent with the original depth, but respectively Set to 64, the total of the four is still 256, the transformation process is shown in Figure 3.
(4) Combine four convolution kernels of different sizes together. Use four different convolution kernels for multi-channel convolution, replacing the third convolution layer of the classic AlexNet network. Compared with a single size convolution kernel, it strengthens the diversity of convolution kernels, and at the same time, it can extract more features.

### 3.3. Normalization

The current network is getting deeper and deeper and the training is more complicated. If the first few layers of the network change, this small change will accumulate and be amplified, that is, the data distribution will change, and it is necessary to relearn this data distribution. If the distribution of training data has been changing during the training process, it will affect the training speed of the network. Since the feature values extracted by the convolutional layer are positive and negative, the negative features will be discarded by the ReLU function, resulting in a certain offset of the feature value after the ReLU activation function. This offset will be affected by the deepening of the network layer. Constantly zooming in will eventually affect the network to make correct judgments. Therefore, choose to set the BN layer between the convolutional layer and the activation function, and use the BN algorithm to normalize the characteristics of the input activation function to reduce the amount of offset caused, thereby accelerating the network convergence and improving the generalization of the network ability [10].

The batch normalization algorithm (BN for short) can reduce the data offset caused by the ReLU function, and solve the problem of the original data distribution changing during the training process, thereby accelerating network convergence and improving network performance. The BN algorithm realizes feature normalization by calculating the mean and variance in a mini-batch. The following briefly introduces the BN algorithm.

The BN algorithm only normalizes the data with a mean value of 0 and a variance of 1. Mainly carry out a two-step operation on the data. The first step: the data is normalized, so that the new data has zero mean and unit variance, as shown in the following formula (1):

\[
\hat{x}^{(k)} = \frac{x^{(k)} - E(x^{(k)})}{\sqrt{\text{Var}[x^{(k)}]}} \tag{1}
\]

Among them, the neuron regarded as the input is the average value of the training data neuron; and the denominator is a standard deviation of the activation of the training data neuron. If a layer only uses this normalization formula to reduce its expressive ability, it will affect the network learning characteristics. For this reason, it is enough to introduce the learnable parameters \(\gamma\) and \(\beta\) to each neuron. The formula is shown in (2):

\[
y^{(k)} = \gamma^{(k)}\hat{x}^{(k)} + \beta^{(k)} \tag{2}
\]

The calculation of the mean and variance used in the above formula is for the entire data set, so the calculation will be too complicated. You can use a simplified way to replace the mean and variance of...
the entire data set with the mean and variance of a batch. This greatly reduces the amount of
calculation. The calculation formulas for the mean and variance of a batch are shown in formulas (3)
and (4):

\[ \mu = \frac{1}{m} \sum_{i=1}^{m} x_i \]  

\[ \sigma^2 = \frac{1}{m} \sum_{i=1}^{m} (x_i - \mu)^2 \]  

Where \( m \) is a batch-size, that is, the size of a batch. With the mean and variance, then ordinary
normalization operations can be performed. The formula is shown in (5):

\[ x_i = \frac{x_i - \mu}{\sqrt{\sigma^2 + \varepsilon}} \]  

According to formula (2) and formula (5), we get the output formula of batch normalization as
shown in (6):

\[ y_i = x_i \gamma_i + \beta \]  

4. Experimental part

The operating system used in this experiment is the Ubuntu 14. 04 platform, the programming
environment is based on Python, the graphics card is RTX3080, and the experiment is carried out
under the deep learning framework Caffe. Use the improved model proposed in this paper to train on
the data set; finally, use the trained model to test on the data set and compare it with the network
method before the improvement to prove the effectiveness of the proposed method.

The public data set SIXray used for two classifications has 1059231 X-ray images. Due to the
excessive number of contraband classifications, this article uses two representative contrabands:
knives and guns to complete the experimental test. On the basis of the SIXray data set, 5, 000 images
containing all the contraband of knives and guns were selected and manually labeled by labeling,
which constituted the data set used for network training in this experiment. The marked information
includes the category and location of the contraband. The coordinates of the upper left corner and
the lower right corner of the marked box are used to indicate the position of the contraband. The crossover
ratio of the output detection box and the marked box is used to determine the matching score to
determine the accuracy. Part of the training set samples are shown in Figure 4.1.

Figure 4. Structural drawing after transformation.

Figure 5. Training set sample image.
When training the network, the influence of parameters such as batch size and learning rate need to be considered. The batch size is the random sample size used in each gradient optimization, which is related to the training time of the network and the recognition accuracy. The size of the learning rate determines the update range of the network parameters. In the initial stage of network training, a larger learning rate helps the network to converge quickly. As the number of iterations increases, the network approaches its best point, and the learning rate can be appropriately reduced to reduce parameter updates. Amplitude, so that the network converges to the best point.

Three batch sizes of 30, 60, 120 are selected respectively, and the relationship between them and the recognition accuracy is analyzed. Table 1 compares and analyzes the performance of the X-ray dangerous goods identification model when the batch size is 30, 60, and 120. It can be seen that when the batch size is 60, the test accuracy of the tool and the gun can reach the best final convergence accuracy.

| base learn | batch size | mAP (Knives) | mAP (firearms) |
|------------|------------|--------------|----------------|
| 0.001      | 30         | 0.7524       | 0.7422         |
| 0.001      | 60         | 0.8431       | 0.8523         |
| 0.001      | 120        | 0.7852       | 0.7562         |

In order to verify the superiority of the improved AlexNet network in the X-ray contraband detection performance, the classic original AlexNet network and the improved AlexNet network were compared experimentally under the same equipment and working environment. The model measurement index was the average precision average MAP (Mean Average Precision) said that Table 1 shows the detection accuracy of the original AlexNet network and the improved AlexNet network on the SIXray data set.

| base learn | batch size | mAP (Knives) | mAP (firearms) |
|------------|------------|--------------|----------------|
| Original AlexNet network | 0.001 | 60 | 0.7672 | 0.7326 |
| Improved AlexNet network | 0.001 | 60 | 0.8525 | 0.8486 |

It can be seen from the experimental results that the improved AlexNet network’s detection accuracy for cutting tools is 8.53% higher than that of the original AlexNet network, and the detection accuracy for guns is 11.6% higher. The significantly improved AlexNet network has better detection performance, which greatly improves the detection rate of hazardous materials in X-ray images.

5. Conclusion
With the development of science and technology and neural network applications, it is a general trend in the future to replace traditional artificial methods for the identification of dangerous goods in X-ray images with automatic identification and detection tools such as neural networks. This article uses the AlexNet network as the basic tool, combined with practical problems and the understanding of the defects of the classic AlexNet network, to improve the classic AlexNet network. Multi-channel convolution technology and normalization are added. The size of the convolution kernel of the convolution layer in the classic AlexNet network is a single size. The generated feature map features are not diverse, and the second is replaced by multi-channel convolution. A single-size convolution kernel in a convolutional layer; the classic AlexNet network has a deeper number of layers, with 11 layers. If a small change occurs in the first few layers of the network, the small change will be accumulated after reaching the back. If you zoom in, the data distribution will change and the gradient explosion will occur. In order to prevent the gradient explosion caused by the different distribution of
the data, this article adds the BN layer, and the convergence speed of the Alex Net network after adding the BN layer is faster, and the final effect is slightly better. Finally, this point is confirmed by comparing the experimental results. For the identification test results of dangerous goods in X-ray images, the improved Alex Net network's test accuracy for cutting tools is 8.53% higher than the training accuracy of the original Alex Net network. Compared with the original Alex Net network, its detection accuracy is also improved by 11.6%. Compared with the traditional identification of dangerous goods in X-ray images, the accuracy and difficulty of detection are improved.

References

[1] Liang Tianfen. A review of research progress on X-ray image detection technology for contraband[D]. Guangdong University of Technology, 2021.

[2] Han Ning. Research on X-ray image dangerous goods detection and tracking algorithm based on deep learning[D]. Lanzhou University, 2018.

[3] Gu Lexu. Tool recognition system design based on X-ray security inspection images[D]. Northeast Dianli University, 2018.

[4] Mark Everingham, SM Ali Eslami, Luc Gool, Christopher K I Williams, John Winn, Andrew Zisserman. The Pascal Visual Object Classes Challenge: A Retrospective [J] . International Journal of Computer Vision. 2015, (1).

[5] Xu X, Tao Z, Ming W, et al. Intelligent monitoring and diagnostics using a novel integrated model based on deep learning and multi-sensor feature fusion[J]. Measurement, 2020, 108086: 0263-2241.

[6] Zhang Jicun. X-ray image intelligent review system based on deep learning[D]. Dalian Jiaotong University, 2021.0125-06.

[7] Huang Lizhi. Research on target detection algorithm based on deep convolutional neural network[D]. Southwest Jiaotong University, 2018.

[8] Song Xiuzhu. Research on X-ray security inspection image processing method [J/OL]. Shenyang Ligong University, 2013 (11): 1-2 [2013-10-22].

[9] Cheng Xin. Research on image target location and recognition based on deep learning[D]. University of Electronic Science and Technology of China, 2016.

[10] Dong Hao. Research on X-ray image analysis method based on multi-view[D]. Nanjing: Southeast University, 2016: 38-49.