Computer Science and Information Systems 18(2):499–516 https://doi.org/10.2298/CSIS200316007S

Image Target Detection Algorithm Compression and Pruning Based on Neural Network

Yan Sun[a] and Zheping Yan[b,*]

College of Automation, Harbin Engineering University,
Harbin 150001, China
{[a] aidenby, [b] feiyunsy1213}@163.com

Abstract. The main purpose of target detection is to identify and locate targets from still images or video sequences. It is one of the key tasks in the field of computer vision. With the continuous breakthrough of deep machine learning technology, especially the convolutional neural network model shows strong Ability to extract image feature in the field of digital image processing. Although the model research of target detection based on convolutional neural network is developing rapidly, but there are still some problems in practical applications. For example, a large number of parameters requires high storage and computational costs in detected model. Therefore, this paper optimizes and compresses some algorithms by using early image detection algorithms and image detection algorithms based on convolutional neural networks. After training and learning, there will appear forward propagation mode in the application of CNN network model, providing the model for image feature extraction, integration processing and feature mapping. The use of back propagation makes the CNN network model have the ability to optimize learning and compressed algorithm. Then research discuss the Faster-RCNN algorithm and the YOLO algorithm. Aiming at the problem of the candidate frame is not significant which extracted in the Faster-RCNN algorithm, a target detection model based on the Significant area recommendation network is proposed. The weight of the feature map is calculated by the model, which enhances the saliency of the feature and reduces the background interference. Experiments show that the image detection algorithm based on compressed neural network image has certain feasibility.

Key words: Convolutional Neural Network, Target Retrieval, Deep Learning, Algorithm Compression

1. Introduction

Images are an important way for humans to access information [1]. With the development of science and technology, image generation is getting faster and faster [2]. Computer image recognition plays an important role in many industries and is a hot topic in current research. Traditional image recognition technology uses artificial feature selection, pattern matching, linear classification and other algorithms for image recognition. The accuracy of the identification depends to a large extent on the quality of the selected features. It is difficult to extract features that express the nature of the

*Corresponding author

original data. Compared with traditional image recognition methods, deep learning has great advantages [3]. A common method is to construct a deep neural network and train it with certain data [4]. The deep neural network can automatically extract image features and achieve better recognition results. Convolutional neural networks are the key network for deep learning and image recognition [5]. By constructing a convolutional neural network, image features can be extracted layer by layer. Network construction and training methods are the key to neural network recognition. Excellent network design can achieve better training results with fewer parameters. Special network components speed up the training process. Appropriate training methods can leverage the capabilities of the network [6].

However, due to the complexity of the real world background and the diversity of scenes, as well as the occlusion and low resolution of the targets in the acquired images, the target detection technology becomes a challenging subject [7]. The main challenges facing current target detection technologies include how to reduce the impact of target size and shape on detection, how to improve the accuracy of target positioning, and how to reduce background interference. The commonly used evaluation indicators for target detection systems are detection accuracy and speed. In order to improve the detection accuracy, the target detection system needs to be able to effectively eliminate the interference of background, light, noise and other factors [8]. In order to improve the detection speed and realize real-time target detection, the target detection system needs to be able to simplify the detection process and image processing algorithms. Since the traditional target detection algorithm is manually designed [9], its accuracy cannot be adapted to various scenes, and the detection speed is slow. Considering the current research status and technical level at home and abroad, this paper will use convolutional neural network to perform object compression in algorithmic compression and pruning for deep learning [10].

Yang et al. extended the clutter model from complex feature vector to complex feature subspace, which is suitable for non-uniform patching regions, and derives extended PTD and extended GP-PNF [11]. Yang et al. proposed a novel supervised target detection algorithm that uses a single target spectrum as a priori knowledge. His proposed algorithm uses TV to maintain the spatial uniformity or smoothness of the detected output. At the same time, the constraints are used to guarantee the spectral characteristics of the unsuppressed target [12]. The final test model is the l1 norm convex optimization problem. The split Bregman algorithm is used to solve the optimization problem because it can effectively solve the l1 norm optimization problem, two synthesis and two real hyperspectral images are used for experiments [13]. Zhao et al. proposed a new hyperspectral image (HSI) target detection method, which uses St OMP reconstruction algorithm [14]. When the computational cost of the conventional sparse detection algorithm is very high, since the HSI usually has a large amount of data, St OMP can be used. The sparse representation algorithm has been successfully applied to the HSI target detection field and has achieved good results. The method improves the steps of solving the sparse coefficient, reduces the number of iterations of the process, significantly improves the detection efficiency and reduces the computational cost [15].

This paper first studies the target detection of convolutional neural networks. Its most widely used target detection deep learning model is constructed by the human visual system. Each layer of the convolutional neural network is described in detail. Of course, the convolutional neural network mainly includes forward propagation and back

propagation. After training and learning, the forward propagation mode will appear in the application of CNN network model, providing image feature extraction, integration processing and feature mapping for the model. The use of back propagation makes the CNN network model have the ability to optimize learning and The algorithm is compressed. Then the research on the Faster-RCNN algorithm and the YOLO algorithm is discussed. Aiming at the problem that the candidate frame extracted in the Faster-RCNN algorithm is not significant, a target detection model based on the attention area recommendation network is proposed. By paying attention to the model to calculate the weight of the feature map, the saliency of the feature is enhanced and the background interference is weakened. Experiments show that the image detection algorithm based on compressed neural network image has certain feasibility.

2. Proposed Method

2.1. Target Detection Theory Based on Convolutional Neural Network

Convolutional neural network

Convolutional neural network is the most widely used target detection deep learning model constructed by human visual system. Convolutional neural networks are a deep learning model of multi-layer neuron connections. Different features can be obtained by convoluting the input images of different convolution kernels. Convolution can be input directly from the original image without complex image preprocessing, so it is widely used. The common layer structure of the convolutional neural network includes a convolutional layer, a pooled layer, a fully connected layer, an activation function layer, and a classification layer. The structure and function of each layer are described in detail below.

As the main structure of extracting image features in CNN network model, convolutional layer plays an important role in the success of CNN network model. The sparse connection mechanism and sharing scheme in the convolutional layer can control the number of parameters and calculations of the convolutional layer to an acceptable range. It is assumed that the L layer is a convolutional layer and the L-1 layer is a pooling layer or an input layer. Then the formula for calculating the level L is (1):

$$x_j^l = f\left(\sum_{i \in Mj} x_i^{l-1} * k_{ij}^l + b_j^l\right) \tag{1}$$

Wherein, the j-th feature map of the L-th layer of the x_j^l th type, the right side of the formula (1) is a convolution operation for all the associated feature maps x_j^{l-1} of the L-1 layer and the j-th convolution kernel k_{ij}^l of the L-th layer And, plus an offset parameter, and finally enter the excitation function.

The sampling layer, also known as the pooling layer, is an important part of the CNN network model, which can reduce the resolution of image features and the computational complexity of the model. In general, it appears after the convolutional layer to perform statistical operations on the small-scale image features extracted from the previous convolution. The sampling layer is static and its parameters do not involve correction of back propagation. There are two common sampling layers: maximum and average. Common pool operations include average pools and maximum pools. As the name suggests, the mean pool selects the average of all values in the slice as the output of the slice, while the maximum pool selects the maximum value of all values in the slice as the output of the slice. For example, maximizing the output of the previous convolutional layer retains only 25% of the activation information, which is the most commonly used pooling method in most convolutional neural networks.

The fully connected layer is usually located at the top of the entire CNN network model as a classifier for the entire model. The input of each node of the fully connected layer is connected with each output node of the previous layer, and the image features extracted by the previous one layer and the pooled layer are mapped to the label space of the sample. Due to the large number of parameters in the CNN network model, over-fitting is caused when training the CNN network model, resulting in insufficient robustness of the model. In order to avoid overtraining and fitting of the CNN network model, the Dropout operation is usually performed at the fully connected layer.

The main role of the activation layer is to give the model a nonlinear fit. In image processing, the convolution operation actually assigns a weight to each pixel in the input layer. Obviously, this is a linear model, but the high-level semantic information of the image is mostly nonlinear, so the model should introduce the ability to fit nonlinearity [16]. This is the case. There are many activation functions, such as sigmoid (Figure 1 from the network, *http://www.baidu.com*), hyperbolic tangent, rectified linear unit, and parametric rectification linear unit (PRELU).

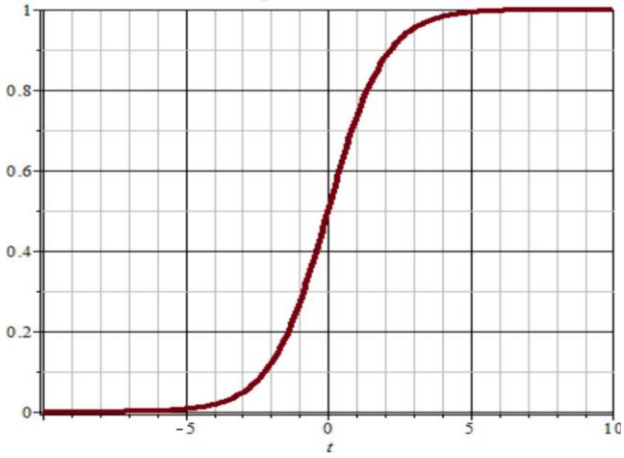

Fig. 1. Sigmoid function

The sigmoid function is a very common function in statistical learning. Its main function is that the model is relatively simple and the algorithm is easy to learn. In machine learning, for large-scale learning tasks, generally better results can be achieved. Sigmoid has also been widely used as an activation function in neural networks. The

advantage is that the derivation is simple and stable, and generally good results can be obtained. However, due to the asymmetry Sigmoid function, function value with respect to the origin is always greater than zero, the average of the output value is always greater than zero, which will reduce the speed of neural network training [17].

The ReLU activation function, as shown in (2), was proposed by Krizhevsky A. It is very popular in neural networks. The expression of the activation function in the network is sparser than the s-type function. Under a random gradient, ReLU decreases rapidly. The s-type function contains many exponential operations, which increases the computational complexity of network training. ReLU can be simpler. Most importantly, this activation function effectively alleviates the gradient dispersion problem of the s-type function. Surprisingly, it also performs well under unsupervised training.

$$y = \max(0, x) = \begin{cases} 0 & x \le 0 \\ x & x > 0 \end{cases} \tag{2}$$

The primary function of the classification layer is to map the output of the entire connection layer to the probability that the input image belongs to a different category. The Logistic function and the Softmax function are the most commonly used classification functions in the classification layer [18]. The Logistic classification function is based on the Bernoulli distribution. It is suitable for binary classification problems. Its function expression is as follows (3).

$$\sigma(z) = \frac{1}{1 + e^{-z}} \tag{3}$$

Considering its output as the probability of t=1, t=0, you can get (4) and (5):

$$P(t = 1 \mid z) = \sigma(z) = \frac{1}{1 + e^{-z}} \tag{4}$$

$$P(t = 0 \mid z) = 1 - \sigma(z) = \frac{e^{-z}}{1 + e^{-z}} \tag{5}$$

The Softmax classification function is mainly used to solve multi-class classification problems. If the dimension of the input vector Z C will be a Softmax classification function, where C is the classification of the category and the output of the quantity is the dimension of a vector, but the elements of the probability vector at this time represent the input image, belonging to each category (6). When the denominator fill uses a regular term, the range of all outputs is limited to (0, 1).

$$y_c = \upsilon(Z)_c = \frac{e^{Z_c}}{\sum_{d=1}^{c} e^{Z_d}} \tag{6}$$

Compression of target detection algorithm based on convolutional neural network

Convolutional neural networks mainly include forward and backward propagation (BP). After training and learning, the application of CNN network model will appear in forward propagation mode, providing image feature extraction, comprehensive processing and feature mapping for the model [20]. The layers of the entire CNN network model and their parameters will participate in the forward propagation. Forward propagation is the main mode of propagation of the CNN network model. Forward propagation only occurs during the training and learning phases of the CNN network model. The parameters of the CNN network model are adjusted layer by layer using a correlation algorithm such as stochastic gradient descent (SGD), so that the output of the loss function in the CNN network model is gradually reduced. The use of backpropagation enables the CNN network model to have the ability to optimize learning. Backpropagation is an indispensable way of spreading the CNN network model. The function can be expressed as J (W, b), and the mathematical expression is as follows (7):

$$J(W, b) = \frac{1}{n}\sum_{i=1}^{n} J(W, b, x^i, y^i) = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{2}\left|\left| h_{w,b}(x^i) - y^i \right|\right|^2\right) \quad (7)$$

Where W and b are the weights and offsets of the convolutional neural network, and the equations for updating the weight parameters and the bias terms are expressed as follows (8) (9) , where W_{ij}^{l} represents the i-th in the first layer in the convolutional neural network model. Enter the weight to connect to the Jth neuron. b_i^l denotes an offset term of the first input connected to the i-th neuron in the first layer in the convolutional neural network model.

$$W_{ij}^{l} = W_{ij}^{l} - \alpha \frac{\partial}{\partial W_{ij}^{l}} J(W, b) \quad (8)$$

$$b_i^l = b_i^l - \alpha \frac{\partial}{\partial b_i^l} J(W, b) \quad (9)$$

2.2. Faster-RCNN Detection Algorithm

Faster-RCNN algorithm

The fast RCNN algorithm implements the real end-to-end target detection calculation process [19], which is mainly divided into three parts: convolutional neural network; regional recommendation network (RPN); Fast R-CNN target detection network. The algorithm still continues the idea that R-CNN first recommends regional reclassification. However, it successfully implements the task of using a convolutional neural network to recommend regions without using other algorithms. RPN and FastR-

CNN share the convolutional neural network for feature extraction, which reduces the number of convolution calculations and improves the speed of the whole algorithm. The R-CNN algorithm requires the extraction of features from each of the recommended candidate regions during training and testing, which requires a lot of practice and is expensive both in time and space. An important reason for these problems is that there is no shared convolution calculation [21].

R-CNN algorithm convolving all input image region, but it must be present in the overlapping part candidate region. Thus, this would create a convolution calculation of the overlap region, increasing the computational complexity of the overall testing process. Then, the FastR-CNN algorithm improved these issues. Although they all use a selective search algorithm to calculate candidate regions, the locations of convolution calculations are different. FastR-CNN first calculates the convolution of the whole image, then fuses the recommended candidate region selection search algorithm and convolution calculation feature mapping network, and obtains the candidate region of the corresponding feature vector through the RoIPooling layer, which greatly reduces the convolution calculation of operation sharing. . Reduced the number of convolution calculations. The dimensionality of these feature vectors is uniform, which facilitates subsequent classification work. FastR-CNN is inspired by SPP-Net. The proposed pooling layer pools the convolution feature and the candidate region boundary to obtain the feature vector of the corresponding region. This is equivalent to a special SP network space pyramid pool with only one pyramid structure. This is the case. In addition, in order to achieve better training results, FastR-CNN also uses some methods to accelerate, two of the most important methods are: multi-task training and minimum batch sampling.

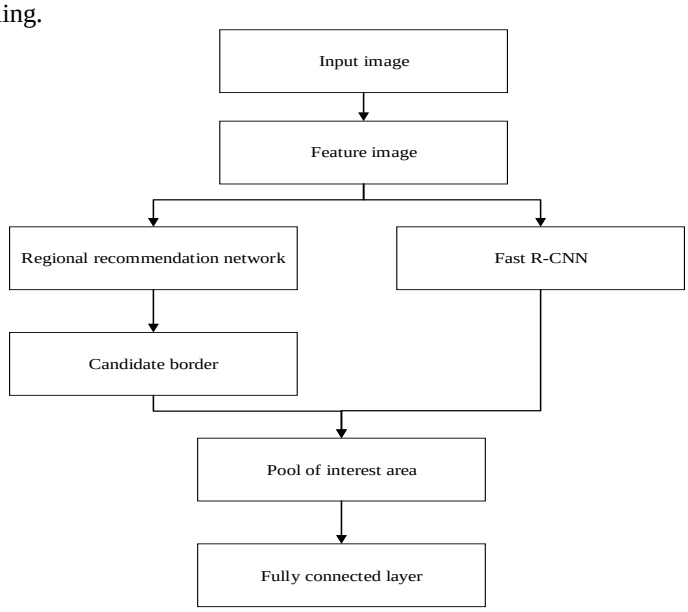

Fig. 2. Faster R-CNN algorithm flow chart

Based on the previous experience of R-CNN and FastR-CNN, the convolutional network is further used to implement the regional recommendation process. The FasterR-CNN network model does not require additional algorithms to recommend

regions. It is an end-to-end complete network model. In the FasterR-CNN algorithm, the input image is directly sent to the convolutional neural network, the feature image of the image is calculated, and then the final convolved feature image is sent to the RPN network of the regional recommendation network to recommend the candidate region. The recommended candidate region frames and their corresponding feature regions are then passed through the FastR-CNN algorithm. The proposed RoIPooling layer is merged into a feature vector of a fixed dimension. Finally, like the FastR-CNN algorithm, the feature vector is input into the classifier and the boundary regression calculator for parallel classification and boundary regression correction. The FasterR-CNN network model integrates the regional recommendation process. The RPN network shares the convolution calculation with FastR-CNN, calculates the image convolution characteristics, and forms an end-to-end target detection model, as shown in Figure 2.

It can be seen that the three algorithms of the R-CNN series are getting faster and better. One of the important reasons is the degree of sharing of convolution calculations. Representative algorithms for primary image detection based on convolutional neural networks are YOLO and SSD. The main features of the algorithm are fast, simple, real-time, but its accuracy is far lower than the R-CNN series. YOLO is one of the representatives of the first-level image detection algorithm. There are three main innovations. The first is to train end-to-end networks. Secondly, the extraction method of the regional recommendation box is improved. The third is to achieve real-time detection results. The YOLO image detection algorithm consists of 24 concave and convex faces and 2 fclayers. This is the case. The difference between YOLO and RCNN, Fast-RCNN and FasterRCNN is reflected in two aspects. The first aspect is that YOLO does not extract the region candidate box steps, saving the detection time of the model algorithm. The second aspect is to obtain the location and type of all objects in the input image and its confidence through the YOLO framework. But the shortcomings are also obvious. There are four largest pool layers in the network framework, so the image features learned through the feature extraction network are not detailed enough, which has a great impact on the accuracy of small target detection.

The YOLO detection model divides the input image into S*S grids, each of which is responsible for predicting the target information of the center. Each grid prediction target belongs to a certain class B bounding box information and C probability. The information of the bounding box includes the spatial position (x, y) and size (w, h) of the bounding box and the confidence level of the bounding box. Confidence is used to reflect whether the current border contains objects, and if there is an object, the degree of confidence is accurate. In order to transform the target detection task into a regression problem, the YOLO detection model has a vector design loss function with a special network output format, as shown in equation (9):

$$
\begin{aligned}
L &= \lambda_1 \sum_{i=0}^{S^2} \sum_{j=0}^{B} \prod_{ij}^{obj} [(x_i - x_i')^2 + (y_i - y_i')^2 \\
&+ \lambda_2 \sum_{i=0}^{S^2} \sum_{j=0}^{B} \prod_{ij}^{obj} [(\sqrt{w_i} - \sqrt{w_i'})^2 + (\sqrt{h_i} - \sqrt{h_i'})^2] \\
&+ \sum_{i=0}^{S^2} \sum_{j=0}^{B} \prod_{ij}^{obj} (c_i - c_i')^2 \\
&+ \lambda_3 \sum_{i=0}^{S^2} \sum_{j=0}^{B} \prod_{ij}^{obj} (c_i - c_i')^2
\end{aligned}
\tag{9}
$$

Optimized compression based on YOLO detection algorithm

The addition of batch normalization not only improves the convergence of the YOLOv2 model in training and learning, but also improves the generalization performance of the detection model. The high-resolution image is used to optimize the basic network of the detection model, which improves the detection accuracy of the model. The K-means clustering algorithm is used to process the PASCALVOC dataset. By analyzing the target data in the Microsoft COCO dataset, a more effective target size ratio can be obtained and the precision of the model can be improved. The prediction method of coordinate information in the boundary coordinate system is improved. In the case of not using anchor points, the information of the bounding box can be directly predicted; with fine image features, we can fuse the image features extracted from different layers of the basic CNN network through layers; using multi-scale image training detection model To make the model more adaptable to different sizes of goals; CNN's basic network has also been redesigned.

In the CNN-based image target detection model, the CNN network model is used to extract advanced image features with rich semantic information [22]. However, in order to reduce the computational complexity of the network model and to ensure that the features of the displacement characteristics, expansion and deformation invariance are extracted, the CNN network model will continuously compress the resolution characteristics of the image, which will help to improve the image recognition of the network model. Classification ability. However, the task of image object detection includes spatial location information of the target and category information of the target. Most CNN-based image target detection models use only image features extracted from the high layers of the CNN network model. CNN can extract image features with rich semantic information, but cannot extract image features with sufficient position information at the same time. The image information extracted from the low-level CNN network model may contain sufficient spatial location information of the detected object. At the same time, the image features extracted from the lower layer of the CNN network model have higher resolution than the image features extracted from the higher layers of the CNN network model. Therefore, the image features extracted from the lower layer of CNN are more conducive to detecting relatively small targets in the

image, and improve the positioning accuracy of the detection model [23]. Due to the different features of the low-level and high-level images in the CNN network model, in order to achieve better detection accuracy, the image target detection model needs to fuse the image features extracted by different layers in the model.

Therefore, the YOLOv2 detection model uses the transport layer to fuse the image features extracted from the underlying network model CNN network with the image features extracted from the upper layer of the CNN network. Then the target detection task is predicted based on the fusion feature, which greatly improves the positioning accuracy of the target detection model and the detection accuracy of the small target [24].

2.3. Image Target Detection Algorithm Based on Deep Learning

After determining the network structure, the framework of the model is determined. The function of this model is spatial mapping. In order to determine if the mapped feature is valid, the predicted value needs to be associated with the target value. The objective function is used to relate the model output to the actual output.

The model is assumed to be represented by a function map $f(x \mid \theta)$. Where x represents the given input and θ represents the parameters of the model. The output of the model may or may not be consistent with the actual label y. In general machine learning, the objective function is usually divided into two parts. If the objective function is J, its general expression is (10).

$$J = \frac{1}{N}\sum_{i=1}^{N} L(Y_i, f(x_i \mid \theta)) + \lambda R(\theta) \tag{10}$$

The first is the loss function and the second is the regularization or penalty. The objective function determines the optimal form of the model parameters. However, how to adjust the parameters of the model to make it better and faster to approximate the ideal mapping function requires a good parameter optimization algorithm. The optimization method can be understood as a model parameter learning algorithm. First, an overview of the basic concepts of optimization is outlined. To simplify the description, the function is represented by F(x), which is the process of adjusting the function F(x) by minimizing or maximizing the value of the argument. The form of the optimization problem can be represented by minimization, since the minimization function can be converted by taking the opposite number of maximum functions. The general minimization problem can be described by a formula:

$$x^* = \arg\min f(x) \tag{11}$$

When both the independent variable and the dependent variable are real numbers, the derivative of the function is denoted as F (x). When the derivative of a point is 0, it does not provide information about the direction of the function around it. The point where F(x) = 0 is called the critical point. The global minimum point is the point at which the function has a minimum value in all defined fields, and the necessary condition is F(x) =0. The local minimum point is the point with the smallest value in the small range

around the function. It is characterized by a function point that moves above any minimum step size is greater than it. A sufficient condition for a function to have an optimal solution is that the function has convexity.

The derivative reflects the slope of the function at point x. If the value of the argument changes slightly along the slope direction, the function value of the same size as the derivative is changed. Therefore, the nature of the derivative can well control the falling point of the function, which is the principle basis of the gradient descent. The formula for updating the independent variable along the gradient direction is:

$$x' = x - \eta \Delta x f(x) \tag{12}$$

Equation (12) is the basis of the gradient descent method or the steepest descent method. It should be noted that the above function approximation method is based on only one step. In fact, there are many multi-order approximation methods based on Taylor expansion functions. In addition, the reader should note that the following optimization methods are generally based on multidimensional input for ease of reading, but this article does not distinguish the dimension of the variable.

3. Experiments

3.1. Data Set

Currently, DCNN-based target detection training usually begins with pre-training of the network designed on ImageNet, and then fine-tunes and tests the PASCAL VOC. ImageNet Dataset ImageNet is a dataset created by the Li Fei team at Stanford University for computer vision projects [25]. The image dataset is based on the hierarchical structure of the word network. Each level of word nodes corresponds to hundreds of pictures. Currently, the average number of images per node is 500. ImageNet is currently the largest image data acquisition system. At present, there are 141997122 pictures with 2141 semantic nodes. Each picture is hand-labeled and has a certain quality guarantee. ImageNet has 103, 4,908 images for target detection tasks, 1,000 images with SIFT features, and about 1.2 million images. With the gradual expansion of the data set, ImageNet also has a corresponding image recognition algorithm to compete. The PASCAL VOC dataset will host an annual algorithm competition from 2005 to 2012. The competition provides a standardized set of excellent data sets for image recognition, object detection and image segmentation. In 2007, the database category was fixed to 20 categories. All images in the dataset have labels for classification, identification, and detection, and only some of the data have segmentation labels. This type of dataset typically uses two versions of 2007 and 2012. There are 963 pictures of VOC 2007, including 5011 training/validation groups and 4952 test groups. A total of 24,640 objects were marked. The VOC2012 dataset differs in the detection and segmentation task tags. For inspection missions, VOC2012 includes all the images from 2008-2011. The training/verification set contains 11,540 images, marking 27,450 objects. For the split task, the training/verification set contains all the images from 2007 to 2011, with 2913 images, marking 6929 objects, and the test set is only 2008. By 2011, because the 2007 test dataset did not release tags, only images.

3.2. Experimental Platform

The hardware platform processor used in this lab is Intel (R) Core (TM) i7-6700@3.40GHz with 16.0GB RAM memory and NVIDIA GeForce GTX1080 GPU graphics card. Among them, the graphics card can provide parallel and accelerated computing for the CNN network model on a reasonable software platform [26]. The operating system used in the experiment was Ubuntu 16.04. The software platforms used are Caffeine 2 and OpenCV 3.2. The programming language used is C++, Python.

4. Discussion

4.1. Analysis of Image Retrieval Results

Fig. 3 shows the target boundary results detected before and after adding a location network. The first is the original map, the second is the result of the frame obtained by the FasterR-CNN algorithm, and the third is the target frame detected after optimizing the network with the target location. In the first two figures, it can be seen that the position optimization network can help the boundary to more closely surround the detection target. The third group of improved maps can be seen that the position optimization network can also completely surround the target. The initial test only revolved around the bird's body, ignoring the bird's tail. After position optimization, the boundary covers the tail of the bird, closer to the label result of the data set. It is effective to expand the search area of the original candidate frame to a certain multiple.

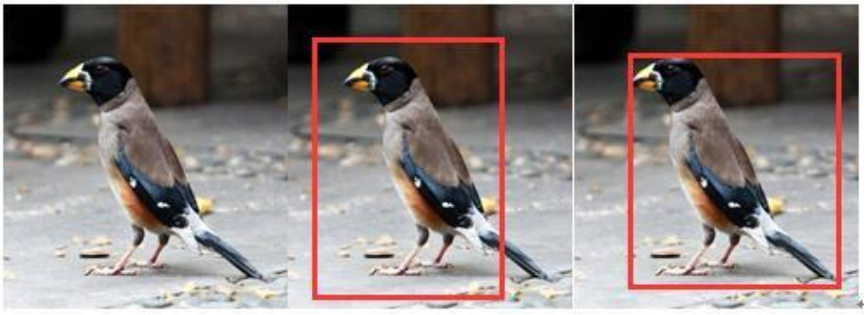

Fig. 3. Comparison of image retrieval results

4.2. Image Detection Quantity Index Comparison

Image detection algorithms based on candidate regions, including R-CNN, Fast-RCNN, and FasterRCNN, are excellent indicators for detecting image accuracy. Another indicator of image detection is the time required to detect an image. Obviously, there is still a gap between these algorithms and real-time algorithms. It can be seen from Table 1 and figure 4 that the speed of the SSD is more advantageous in detecting the quantity quality of the picture within 1 s.

Table 1. Image detection quantity comparison

Method	Image detection
RCNN	0.03
Fast RCNN	0.49
Faster RCNN	6
YOLO	46
SSD	60

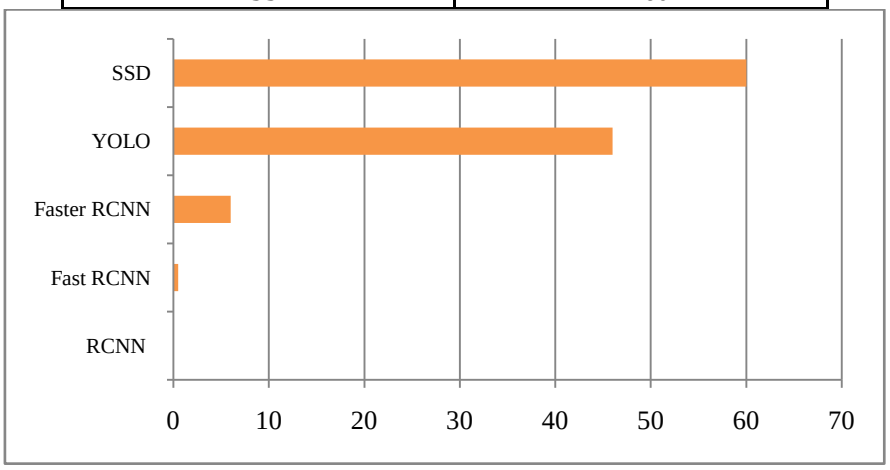

Fig. 4. Image detection quantity comparison

Image target recognition is performed using the FasterR-CNN method. Convolutional neural networks have many options. The public network can be used here as a convolutional neural network, removing the fully connected layer of the network and the pooling layer at the end of the network. The purpose of this method is to reduce the image. On a scale, each region of the image feature map has an area corresponding to the original image.

4.3. Target Detection

The average test accuracy of all image categories from the PASCAL_VOC2012 data set can be seen from Figures 5 and 6. The MFCN algorithm proposed in this paper is superior to other target detection algorithms in mAP. The mAP of MFCN is 73.2%, which is 2.8% and 0.8% higher than Faster_R-CNN and SSD, respectively. In addition, compared with YOLO, the effective fusion of MFCN multi-features and multi-frame prediction greatly improves the detection accuracy of MFCN for small targets such as birds. The above results show that the target detection result generated by the algorithm has higher precision and can better solve the small target detection problem.

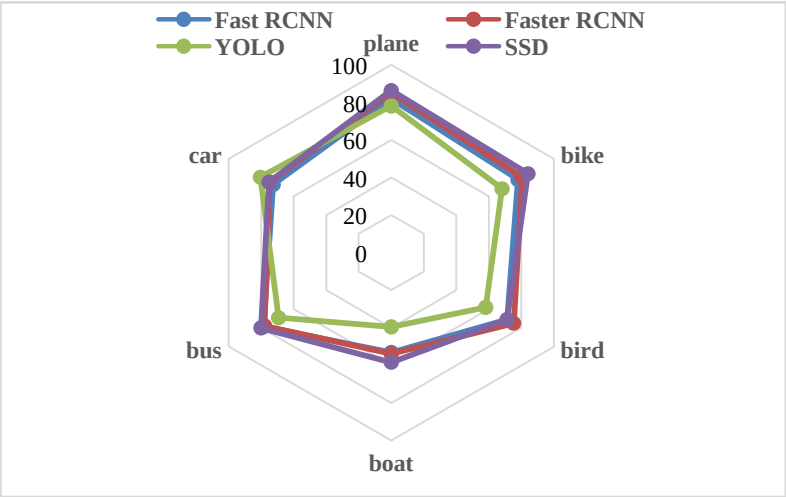

Fig. 5. Image detection results accuracy comparison 1

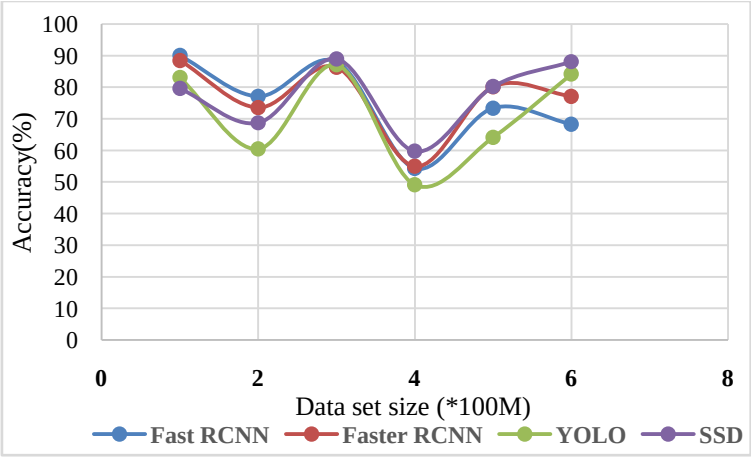

Fig. 6. Image detection results accuracy comparison 2

A target detection framework based on the regression idea of YOLO. Aiming at the problem of YOLO, a target detection method based on full convolutional network and multi-feature fusion is proposed. By establishing a convolutional neural network with no fully connected layers for target detection, since the size of the input image is not limited, the model can be used to detect multi-scale targets and can be trained using multi-scale images. The algorithm also makes full use of the feature information of different depths to obtain the feature information rich in the detected target, and improves the detection accuracy of the small-scale target.

4.4. Comparison of Storage Costs at Different Compression Ratios

Use the method of this paper to evaluate the performance of the weight pruning method on the data set. The pruning is obtained for the ownership weight below the threshold in the network to obtain the compression model. As shown in Figure 7, the storage overhead of this method is compared with different compression ratios. At 90% compression, the maximum storage cost is reduced, and the accuracy is only reduced by 0.21%.

The performance of the convolution kernel pruning method is evaluated on the dataset using a neural network. Starting from the collection of training sets for channel selection, 10 images are randomly selected from each category in the training set to form an evaluation set. And for each input image, 10 instances are randomly sampled with different channels and different spatial locations. Therefore, there are a total of 1000 training samples used to find the best subset of channels through the greedy algorithm. Experiments have demonstrated the effectiveness of this choice (10 images per class, 10 locations per image), sufficient for neuronal importance assessment. Fine tune each layer after pruning. When all layers have been trimmed, fine-tune 10 times for greater accuracy.

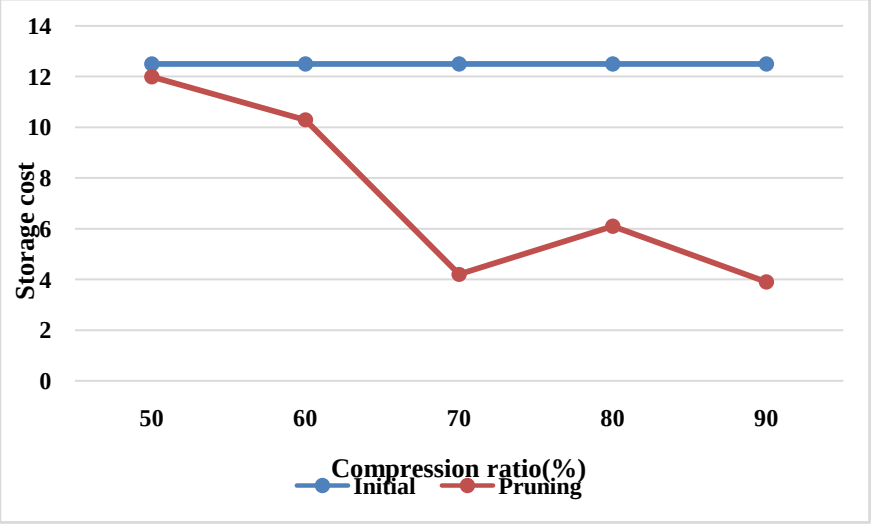

Fig. 7. Comparison of storage costs at different compression ratios

5. Conclusions

(1) There are many redundancy in the parameters of the CNN network model. Firstly, the convolutional neural network target detection is studied. Its most widely used target detection deep learning model is constructed by the human visual system. Each layer of the convolutional neural network is described in detail. Of course, the convolutional neural network mainly includes forward propagation and back propagation. After

training and learning, the forward propagation mode will appear in the application of CNN network model, providing image feature extraction, integration processing and feature mapping for the model. The use of back propagation makes the CNN network model have the ability to optimize learning and the algorithm is compressed.

(2) The Faster-RCNN algorithm and the YOLO algorithm were discussed and discussed. Aiming at the problem that the candidate frame extracted in the Faster-RCNN algorithm is not significant, a target detection model based on the attention area recommendation network is proposed. By paying attention to the model to calculate the weight of the feature map, the saliency of the feature is enhanced and the background interference is weakened. .

(3) Construct and train deep neural networks with certain data. The deep neural network can automatically extract image features and achieve better recognition results. Convolutional neural networks are the key network for deep learning and image recognition. By constructing CNN, image features can be extracted layer by layer. The comparison test results of several algorithms show that the number of parameters of the model is greatly reduced, and the feature representation ability of the model within the acceptable range is reduced.

Acknowledgment. This paper is funded by National Natural Science Foundation of China under Grant No. 51679057

References

1. Zafar, Bushra; Ashraf, Rehan; Ali, Nouman; Ahmed, Mudassar; Jabbar, Sohail; Naseer, Kashif; Ahmad, Awais; Jeon, Gwanggil. Intelligent Image Classification-Based on Spatial Weighted Histograms of Concentric Circles. Computer Science and Information Systems. 15(3). 615-633. (2018).
2. Tung, Frederick, & Mori, Greg. Deep Neural Network Compression by In-Parallel Pruning-Quantization,IEEE Transactions on Pattern Analysis and Machine Intelligence,PP (99), pp. 1-1. (2018)
3. Díaz, G., Macià, H., Valero, V. et al. An Intelligent Transportation System to control air pollution and road traffic in cities integrating CEP and Colored Petri Nets. Neural Comput & Applic 32, 405–426 (2020).
4. Fan, D., Wei, Lu, & Cao, Maoyong. Extraction of Target Region in Lung Immunohistochemical Image Based on Artificial Neural Network,Multimedia Tools & Applications,75(19), pp.1-18. (2016)
5. Y Zhao, H Li, S Wan, A Sekuboyina, X Hu, G Tetteh, M Piraud, B Menze. Knowledge-aided convolutional neural network for small organ segmentation. IEEE journal of biomedical and health informatics, 23(4), pp. 1363-1373, (2019)
6. Ding, X., & Yang, Hong Hong. A Study on the Image Classification Techniques Based on Wavelet Artificial Neural Network Algorithm,Applied Mechanics & Materials,602-605, pp.3512-3514.
7. Lin, G., Wang, J., Yong, F., & Chen, N. Robust Visual Tracking Based on Convolutional Neural Networks and Conformal Predictor,Acta Optica Sinica,37(8), pp. 815003. (2017)
8. Wang, Li, Tang, Jun, & Liao, Qingmin. A Study on Radar Target Detection Based on Deep Neural Networks,IEEE Sensors Letters,3(3), pp.1-4. (2019)
9. Liu, P., Guo, J. M., Wu, C. Y., & Cai, D. Fusion of Deep Learning and Compressed Domain Features for Content-Based Image Retrieval,IEEE Transactions on Image Processing A Publication of the IEEE Signal Processing Society,PP(99), pp.1-1. (2017)

10. Arun, K. S., & Govindan, V. K. A Hybrid Deep Learning Architecture for Latent Topic-Based Image Retrieval,Data Science & Engineering, 3(2), pp. 166-195. (2018)
11. Yang, Dongwen, Du, Lan, Liu, Hongwei, Wang, Yan, & Gu, Mingfei. Extended Geometrical Perturbation Based Detectors for Polsar Image Target Detection in Heterogeneously Patched Regions,IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing,12(1), pp.1-17. (2019)
12. Wang, G., Yao, Y., Chen, Z., & Hu, P. Thermodynamic and optical analyses of a hybrid solar CPV/T system with high solar concentrating uniformity based on spectral beam splitting technology. Energy, 166, 256-266. (2019)
13. Yang, S., & Shi, Zhenwei.Hyperspectral Image Target Detection Improvement Based on Total Variation,IEEE Transactions on Image Processing,25(5), pp.2249-2258. (2016)
14. Liu, Hui; Li, Chenming; Xu, Lizhong. Dimension Reduction and Classification of Hyperspectral Images based on Neural Network Sensitivity Analysis and Multi-instance Learning. Computer Science and Information Systems. 16(2). 443-467.(2019)
15. Zhao, C., Jing, X.,& Li, W. Hyperspectral Image Target Detection Algorithm Based on Stomp Sparse Representation, Harbin Gongcheng Daxue Xuebao/Journal of Harbin Engineering University, 36(7), pp.992-996.(2015)
16. Ali, Munwar, Low Tang Jung, Abdel-Haleem Abdel-Aty, Mustapha Y. Abubakar, Mohamed Elhoseny, and Irfan Ali. Semantic-k-NN Algorithm: An Enhanced Version of Traditional k-NN Algorithm.Expert Systems with Applications: 113374.(2020)
17. Chi-Hua Chen, Fangying Song*, Feng-Jang Hwang, Ling Wu, "A Probability Density Function Generator Based on Neural Networks," Physica A: Statistical Mechanics and its Applications, 541, Article ID 123344, March.(2020)
18. Lakshmanaprabu SK, Mohamed Elhoseny, Shankar Kathiresan, Optimal Tuning of Decentralized Fractional Order PID Controllers for TITO Process using Equivalent Transfer Function, Cognitive Systems Research, Volume 58, December, pp. 292-303.(2019)
19. Zheng Xu, Lin Mei, Zhihan Lv, Chuanping Hu, Xiangfeng Luo, Hui Zhang, Yunhuai Liu。Multi-Modal Description of Public Safety Events Using Surveillance and Social Media. IEEE Trans. Big Data 5(4): 529-539 (2019)
20. Xiong, Q., Zhang, X., Wang, W., & Gu, Y. A Parallel Algorithm Framework for Feature Extraction of EEG Signals on MPI. Computational and Mathematical Methods in Medicine, 2020, 1-10. (2020)
21. Junlong Chen, Xiaomeng Wang & Zhaopeng Chu Capacity Sharing, Product Differentiation and Welfare, Economic Research-Ekonomska Istraživanja, 33:1, 107-123.(2020)
22. Xu Z, Cheng C, Sugumaran V. Big data analytics of crime prevention and control based on image processing upon cloud computing. J Surveill Secur Saf 2020;1:16-33.
23. Ling Wu, Chi-Hua Chen*, Qishan Zhang, "A Mobile Positioning Method Based on Deep Learning Techniques," Electronics, 8(1), Article ID 59, January.(2019)
24. Mu Zhou, Yanmeng Wang, Yiyao Liu, and Zengshan Tian. An Information-theoretic View of WLAN Localization Error Bound in GPS-denied Environment. IEEE Transactions on Vehicular Technology. 68(4): 4089-4093.(2019)
25. de Souza, L.A., Afonso, L.C.S., Ebigbo, A. et al. Learning visual representations with optimum-path forest and its applications to Barrett's esophagus and adenocarcinoma diagnosis. Neural Comput & Applic 32, 759–775 (2020).
26. Wei, P., He, F. & Zou, Y. Content semantic image analysis and storage method based on intelligent computing of machine learning annotation. Neural Comput & Applic 32, 1813–1822 (2020).

Yan Sun received the B.S. degree in electrical information engineering in 2004, and the M.S. degrees in communication and information system, and the Ph.D. degrees in signal and information processing from Harbin Engineering University (HEU), China, in 2012.

His research interests include wireless communications, image processing and machine learning. E-mail: aidenby@163.com

Zheping Yan was born in Longyou, Zhejiang, China, in 1972. He received the B.S. degree in nuclear power equipment, the M.S. degree in special auxiliary device and system in naval architecture and ocean engineering, and the Ph.D. degree in control theory and control engineering from Harbin Engineering University, China, in 1994, 1997, and 2001, respectively. He was the Post-Doctoral Researcher in mechatronic engineering with the Harbin Institute of Technology, China, in 2004.His current research interests include identification of non-linear system, and multi-sensors data fusion and intelligent control. E-mail: feiyunsy1213@163.com

Received: March 16, 2020; Accepted: February 08, 2021.