Research on battlefield target recognition method based on machine learning

Jiang Yuan Zhen¹, Zhou Zheng¹, Lu Wei¹ and Lu Yuan Gen²

¹ CETC No.11 Research Institute, Beijing, China,
² School of information science and engineering, ShenYang Ligong University, Shenyang

Email: gingeryzh721@163.com, luyuangen97@163.com, China

Abstract. All articles must contain an abstract. The abstract text should be formatted using 10 point Times or Times New Roman and indented 25 mm from the left margin. Leave 10 mm space after the abstract before you begin the main text of your article, starting on the same page as the abstract. The abstract should give readers concise information about the content of the article and indicate the main results obtained and conclusions drawn. The abstract is not part of the text and should be complete in itself; no table numbers, figure numbers, references or displayed mathematical expressions should be included. It should be suitable for direct inclusion in abstracting services and should not normally exceed 200 words in a single paragraph. Since contemporary information-retrieval systems rely heavily on the content of titles and abstracts to identify relevant articles in literature searches, great care should be taken in constructing both.

Keywords: Machine learning, Transfer learning, Reuse features, Target recognition

1. Introduction

As one of the important technologies of machine learning, deep learning is obviously superior to traditional technology in the field of artificial learning. However, deep learning relies heavily on data and computing resources to make training effective models rigorous. Deep transfer learning is a very important tool to solve the problem of insufficient data and early training difficulties. In the field of deep learning, by migrating knowledge from source domain to target domain, it can solve many difficulties in the field of machine learning which can not train better performance models because of insufficient data. Deep learning based on network is to reuse part of the network that has been pre-trained in the source domain, including network structure and connection parameters, and migrate network structure and connection parameters to the neural network of target domain. However, there is a certain difference between the target task and the source domain task. In some samples, the features that may be extracted are very similar, while the other parts are quite different. Therefore, this paper introduces a semi supervised learning mechanism to enhance the data of the larger samples. Improving the ability of network to extract features in this part of the sample. Semi supervised transfer learning method can easily solve the problem of overfitting of the network while solving the problem of partial classification inaccuracy. It is easy to bring the over fitting problem caused by data imbalance when weighting or classifying inaccurate samples. In this paper, a semi supervised augmented transfer learning method is proposed. The enhanced transfer learning mainly uses the concept of model fusion.
in machine learning domain to solve the over fitting problem. By integrating multiple model classifier parameters, the variance error caused by semi supervised learning is eliminated.

The battlefield environment is complex and changeable, and the feasibility of relying on a pure algorithm for identification is very low. The deep neural network learns many features during the training process, and it is very important to select the features that are most suitable for the classifier. In recent years, many classic networks that have achieved good results use grouped convolution and features. The way to choose, such as Google's Xception[1] Network, SE-Net by Momenta and ResNeXt by Facebook [2]: A series of advanced network structures such as the network have adopted these ideas more or less. Based on the above two ideas, this article proposes a multiplexed feature selection mechanism that uses deep neural networks to extract and select target features, combining the feature selectivity of the attention mechanism and the feature reuse of grouped convolution calculations to improve the performance of the feature extractor. The advantage of deep convolutional neural networks is that they can extract image features without rotation and translation, and introduce them into the neural network to learn and optimize convolutions. Kernel parameters, compared to a fully connected network with a smaller number of parameters and calculations.

The application of target recognition in battlefields has always been a hot research topic in military field. Identifying types and attributes of targets in battlefield plays a very important role in fighting security risks and understanding battlefield situation. It is undoubtedly tedious to identify targets by traditional methods, and the results of target recognition are related to each module. A large number of modules can easily accumulate errors in the same system. If the error of the front image processing is large, the segmentation module will be introduced. When the error accumulates to the classifier at a time, it may not be able to accurately identify the target. This paper designs a practical application plan based on the semi supervised reinforcement learning algorithm and the target identification method of the multiplexing feature network. Experiments show that the scheme has good accuracy and robustness.

2. Machine learning

2.1. Statistical theory
Statistical learning theory[3] is a very good tool for machine learning models and understanding complex data sets. In recent years, statistical learning theory and computer science theories have developed parallel to each other in parallel, especially in machine learning. The development of this field has produced many famous algorithms, such as reinforcement learning, Decision tree and support vector machine. In the era of big data explosion, statistical learning theory has become very popular. In market, finance and other commercial establishments, statistical learning theory has become an indispensable professional skill for professional and technical personnel.

Support Vector Machines[4] is the most commonly used machine learning algorithm in statistical learning theory and even in industry. Support vector machine is a machine learning algorithm developed on the basis of statistical learning theory. However, unlike other statistical learning methods, statistical learning is usually based on empirical risk minimization. Support vector machine (SVM) is based on structural risk minimization. A SVM model first creates a feature space, which is a finite dimensional vector space, and each dimension represents the characteristics of a special target. The support vector machine is mainly used to solve the support vector of the super plane and minimize the structural risk. Compared with other empirical risk minimization algorithms, the support vector machine has strong robustness.

Support vector machine has a very large advantage that can be applied to practice. It does not simply limit the ability of classification to data in two-dimensional linear space. By adding various nonlinear kernel techniques, the support vector machine can be more flexible to apply to various types of nonlinear decision boundary. And support vector machine does not consume too much computing resources. But support vector machine has a certain disadvantage. If the number of objects in the classification is more than the number of features, the support vector machine may not be effective.
Moreover, the support vector machine is a non probabilistic model, and the decision boundary is not so good to measure whether it is optimal. Softmax function,[5] also be called a normalized exponential function, which is a generalization of logistic regression in the field of multi-classification. The Softmax function can convert a K-dimensional vector containing any real number Z compressed to $\sigma(z)$ in such a way, the sum of each element is 1. The machine learning classifier usually inputs a feature vector corresponding to the posterior probability of the feature category. Since the output result is a probability form, it must be in the range of 0 to 1. Softmax function compresses the result to 0 to 1 output, which is very suitable as a multi-class classifier. Softmax function expression is as follows:

$$P_i = \frac{e^i}{\sum_{j=1}^{n} e^j}$$  (1)

The Softmax function is very stable for numerical calculations. It will not cause destructive overflow and underflow. It is very adaptable to hundreds of categories. In statistics, Softmax function is usually called "multiple logistic regression" function. When Softmax function is classified as a classification problem, it will degenerate into logistic regression function. The Softmax function is usually placed on the last layer of the neural network. The neural network itself has good scale invariance for feature extraction, and the generalization of Softmax function in multi-classification fields has become an important algorithm in the field of image recognition.

2.2. Deep learning theory

In the past few years, deep learning[6] has developed into one of the most important areas in machine learning. It has been applied to various fields, such as image recognition, Natural Language Processing and speech recognition. Deep learning is based on neural networks with two or more than two implicit inputs to analyze input problems through nonlinear transformation or expression learning. A deep neural network is used to decompose a very complex problem into several relatively simple problems. The complex hierarchical structure makes the deep neural network possess strong feature expression ability, and learn weights and deviations from training data to make the analysis results more accurate. Training a better neural network requires not only proper infrastructure, but also training skills. With the development of deep learning, training techniques are changing rapidly, such as Data Augmentation, Weight Initialization and so on. Training neural networks rely heavily on the number of training set data. Compared with traditional feature extractors, neural networks are characterized by invariance of features, and the advantage of traditional classifiers is generalization of classifiers. These two advantages need to be established in the condition of sufficient data. A way to alleviate the problem of insufficient training parameters in neural networks is to enhance data technology. By changing the existing data without changing the essence, new data can be changed. There are geometric transformations, sampling, mirroring and rotation in common data enhancement techniques in computer vision, Displacement and various optical transformations.

Deep neural networks usually have massive parameters, and their loss functions are generally non convex. Therefore, the training of deep neural networks is generally very difficult. In order to quickly converge and avoid vanishing gradient problem, initialization of neural networks is an important prerequisite. Bias items can be initialized to zero. However, weight parameters need to initialize the symmetry of nodes in the same layer. Commonly used initialization methods include Xavier initialization[7], MSRA initialization, normalization[8] and so on.

3. Design of target recognition method based on semi-supervised enhanced transfer learning

3.1. Design of semi-supervised reinforcement learning methods

Deep migration learning[9] is a very important tool to solve the problem of insufficient data and early training. In the field of deep learning, migrating the knowledge of the source domain to the target domain can solve many difficulties in the field of machine learning which can not train better
performance models because of insufficient data. The learning process of transfer learning is shown in Figure 1 below.

![Figure 1. Processing of transfer learning](image)

The deep migration learning based on network will make the target task and source domain task different. In some samples, the features that may be extracted are very similar, while the other part is quite different. For this reason, a semi supervised learning method is proposed, and its learning process is shown in Figure 2.

![Figure 2. Deep transfer learning process based on network](image)

The semi supervised depth migration learning method proposed in this paper has two ways. First, we can enhance data processing for samples with high classification error rate. Image data enhancement is usually classified into two categories: spatial geometric transformation, pixel point transformation, and spatial geometric transformation for image rotation and translation. The transformation and other airspace operations can enable samples to train more neural networks without losing information. The idea of pixel transformation is the principle of image enhancement and denoising. The neural network itself has strong adaptability to Gauss noise and salt and pepper noise. Therefore, the operation of pixel operation can be enhanced by adding noise and adjusting contrast, lightness and brightness.

The second semi supervised learning mechanism adopts the idea of Focal loss function to process the loss function of classifiers. The cross entropy loss of classified error classes uses a greater than 0 and less than 1 scaling factor $\lambda$. The proportion of total loss is increased, and the loss of cross entropy of other classes is adjusted by using the inverse set of scaling factors.

$$L = \sum_i (1 - \lambda) \cdot y \log \hat{y} + \lambda \cdot y_{sp} \log \hat{y}_{sp}$$

(2)

Through the above way, increasing the error rate of the classification error rate to train the neural network can reduce the difference between the source domain and the target domain, and reduce the poor performance of the classifier in a certain category.

Based on semi supervised depth transfer learning, this paper introduces enhanced transfer learning to solve the over fitting problem caused by semi supervised learning, and the over fitting problem caused
by semi supervised learning. The enhanced transfer learning method proposed in this paper is mainly based on the Bagging idea in the idea of Ensemble Learning. At the same time, 5 or more classifiers are trained in the sample data set, and the first five classifier parameters with the highest accuracy are fused to smooth the variance between the classifiers. It is a classifier that can learn the characteristics of the data set corresponding to the class commonality, thus greatly improving the accuracy.

Before classifiers parameters are fused, each classifier needs to be normalized. There will be some dimensional differences between the classifiers trained by sampling. The fusion of parameter distributions with dimensional differences will have a serious impact on the performance of the classifier. So before classifying, all classifications must be normalized. The classifier after fusion can be used as a classifier of the network directly, or a small number of data sets can be used to fine tune the parameter distribution of the classifier. The classifier fused by Bagging can greatly reduce the variance error of the dataset itself. To solve the over fitting phenomenon caused by semi supervised transfer learning, enhanced transfer learning is shown in Figure 3 below.

3.2. Method validation
In this paper, we compare the convergence speed and stationarity of the 14 layers deep convolution neural network before and after the semi supervised depth transfer learning and the semi supervised in-depth migration learning. In the ImageNet data pre-trained VGG-11 network parameters migrating to the experimental network, we train 5 classifier networks respectively, each batch of small 256. Experiments were carried out on the Cifar-10 dataset[10]. The convolution layer of the model used 3 (3) small convolution kernel. In addition, the learning batch Batch was added in the middle of the scroll layer and the activation function, which reduced the influence of data distribution on the weight distribution of the network, thereby reducing the problem of numerical instability in the transmission process of the feature map.

Training data set Cifar-10 on Tensorflow deep learning platform[11] and selecting data with lower accuracy for data enhancement. As shown in Figure 4, the Loss curve without adding semi supervised learning network and adding semi supervised transfer learning is smoothed to 0.6. By using graph 4 (a) Loss value curve, we can see that there is no deep convolution neural network with transfer learning. The Loss value is not stable in the early stage, and the amplitude of swing is relatively large. After training 30 epoch, it reaches 1.68. diagram 4 (b) and adds the network of transfer learning. The convergence speed is relatively fast and the convergence process is relatively stable. The amplitude of Loss value is smaller than that of the network without learning transfer. The network value is relatively stable. In the thirteenth epoch networks, the Loss value is reduced to about 1.62 when the transfer learning network is added, which proves that the network can converge without too many data sets.
Transfer learning not only has a stable effect on the Loss value of the network, but also has a certain effect on the distribution of weights and the distribution of bias values in the network, as shown in Figure 5 (a), that is, the distribution of the bias values of the first layer of the network without migration learning, and the covariance deviation of the internal variables occurs. The value is relatively unstable. For the deep convolution network, very small covariate offsets will cause great gradient changes, which may largely reduce the performance of the network. Figure 6 (b) is the distribution of the bias values of the first layer of the transfer learning network. It can be seen that the distribution of the bias values always follows the normal normal distribution after the transfer learning. The data distribution of standard normal distribution is very beneficial in the network. The data distribution is always stable. There will not be excessive or too small data, which will lead to the loss of information when the data is transmitted. The effect of transfer learning on the network bias values and weights is obvious. As for the backpropagation of the network, it also has a certain numerical stabilization effect. As shown in Figure 7, the gradient distribution change initiated by transfer learning is more stable.
Each sub classifier is a three level DNN network. The network structure can be the same as that of the classification layer in Table 1. After training the classifier, the data set is sorted according to the accuracy of the model on the test set. In this paper, top-1 accuracy is used to sort the model. The accuracy and ranking of the subclassifiers are shown in Table 1.

| mode | A     | B     | C     | D     | E     |
|------|-------|-------|-------|-------|-------|
| Accracy | % | % | % | % | % |
| Sampling | 80% | 80% | 80% | 80% | 80% |

After selecting 5 models with better recognition rate, normalizing the model parameters, and then averaging the model parameters, the parameters after fusion are no longer the original dimension. The model needs to be fine-tuning. The fine-tuning model still uses Cifar-10 data set and trains 200 epoch to stabilize the model. Using the semi supervised enhanced migration model and the loss curve of the model without migration, as shown in Figure 8, the accuracy rate curve changes as shown in Figure 9.

After tuning through some data sets, the loss value of the network is reduced to 0.36 after the semi supervised enhanced transfer learning in 10 types of data sets of Cifar-10. The accuracy of the test set is 88.12% higher than that of the sub classifier and the classifier without migration learning. The migration of VGG-11 basis convolution kernel to the experimental network not only makes the deep convolution neural network stable in the early stage. It also accelerates the convergence speed of the
network, and initializes the parameters of the classifier to enhance the recognition rate of the neural network by enhancing the transfer learning.

Semi-supervised deep migration learning can play a good role in alleviating the data dependency problem of deep learning, and has a good stabilizing effect on the early convergence of the network. Experiments show that the deep convolution network with semi-supervised reinforcement learning is better than the network with no addition.

4. Sections, subsections and subsubsections

4.1. Design of reuse feature network

Convolution neural network (CNN) has been proved to be an efficient model for processing a variety of visual tasks. For each layer, there is a set of filters that can express the local features of multi-channel images, in other words, convolution filters can combine the learned spatial information and channel information in acceptable fields. A deep convolution neural network model with better effect is constructed by stacking the winding layer with nonlinear activation and down sampling, and convolution neural network is a powerful hierarchical image descriptor. SE-Net [12] is the convolutional neural network that won the championship in the 2017 ILSVRC classification contest. It reduced the ImageNet data set to 2.251% top-5 error rate for the first time. The "extrusion and excitation" module proposed by SE-Net can adaptively calibrate multi-channel feature maps, assign a weight to each channel of the feature map, and enhance or weaken channel characteristics by weight. This idea is like Soft Attention, and a weight distribution is selected to filter out the characteristics.

The attention mechanism is a simple one-dimensional vector, usually calculated by the Softmax function through the weighted connection layer. Before the attention mechanism arises, the translation model relies on reading a complete sentence and compressing the sentence to a numerical vector suitable for length. If a sentence with hundreds of words is expressed by a few words or dozens of words, it will inevitably lead to certain information loss and incomplete translation results. However, attention mechanism can alleviate such problems. Attention mechanism is very good in solving the relationship between global and local feature vectors. Attention mechanism is not mysterious and complex. It is just a mathematical formula consisting of fine mathematical symbols and parameter interaction. In the model, attention mechanism can be added to any location. It can help the model to excavate potential and appropriate information and improve the effect of the model.

When extracting multi-channel image features from convolution neural networks, not all local features of each channel are useful. Feature selection can reduce the computation complexity of convolution network and increase the recognition rate of neural network models. The attention mechanism can well extract the feature selectivity of convolution neural network and extract multi-channel features for an image. Many features may be redundant. In this paper, an attentional mechanism neural network model is proposed for feature selection problem. Attention mechanism convolution neural network model chooses depth features from depth and breadth respectively, weakens the recognition of non-affected features and enhances useful local features.

In the se module, a branch is reconstructed to recalibrate the image feature. Suppose a feature $U$. The SE module is first introduced into the Squeeze operation. This operation reduces the feature map from the spatial feature to the channel feature, and reduces the dimension of the feature map to 1 (1 feature descriptor takes the spatial dimension as the dimension). $H \times W$ It compresses the channel descriptor, which describes the global information into the branch network. Then, it activates (Excitation) operations, where a special activation function is used to learn a self threshold mechanism based on channel dependence. This threshold mechanism is reassigned to each channel weight of the feature graph. The specific process is shown in Figure 10 below.
In the SE-Net, the full link network is used to map the weight distribution of attention. The fully connected network can well map the feature map to the weight distribution of attention mechanism, but the whole connection network has poor ability to extract local information of the feature map and occupy most of the network parameters. In this section, convolution computation is proposed instead of full link network for feature selection. The 1 dimension convolution is used to transform the weight distribution after extruding to learn the attention mechanism coefficient to select feature. Feature selection and convolution feature selection are shown in Figure 11.

The standard convolution calculation method is to compute the initial convolution kernel and the traversal of the feature graph. Although this method reduces the parameter and computation amount compared with the traditional fully connected neural network, it still consumes time and computing resources and can not achieve real-time effect. It can not be applied to industrial equipment practically. Such as sparse convolution kernel, fast Fu Liye transform and some other computing methods, but these methods will cause large loss of accuracy or complex instruction set calculation. The idea of group convolution will not greatly reduce the computation accuracy, but it can greatly speed up the computation. The idea of group convolution is first proposed in AlexNet. In AlexNet, convolution operations are divided into two groups, which are calculated on two GPU respectively. Finally, the feature graph is spliced at the back end of the network. Then Google puts forward Inception series network to integrate the volume into more groups, and the convolution kernel sizes of each group are different from each other. In the paper of Facebook ResNeXt network, it is pointed out that the convolution is not only a simple way to reduce the parameters and the compression network, but also the characteristics of Ensemble Learning. The same as Bagging, the convolution of the feature maps is combined together, which can reduce the problem of over fitting to some extent. The core idea...
of ResNeXt's "separation transformation merge" enables the network to achieve the effect of deep network in a relatively small number of layers.

Convolution layer attempts to learn some convolution kernels in visual space from two spatial dimensions - height and width, and a channel dimension. Therefore, a single convolution kernel task needs to simultaneously learn information related to information and channel dimensions related to spatial dimension. While learning two dimension tasks, the information of spatial dimension and channel dimension will inevitably affect each other, so that convolution kernel can not fully learn the information of all dimensions, so the convolution kernel channel convolution is best carried out separately. This ensures that two convolution kernels can learn their own information independently. For example, the Xception network of Google uses 3 (3 space convolution kernel grouping) to learn. Then we use the 1 (1 channel convolution) learning channel information to splice the feature graph after convolution operation, which not only makes good use of the advantage of small amount of parameter computation of convolution, but also increases the interactivity of convolution. At the same time, the space convolution and channel convolution can fully learn the information of space dimension and channel dimension.

By adding feature selection mechanism in the above packet computing, the network can learn better features in the feature extraction stage, and packet computing itself will reduce the number of network parameters. While reducing the quantity, it also loses part of the feature information of the image. In the form of grouping, the selected feature reuse can slow down the phenomenon of feature overselection while guaranteeing the reusability of the feature, and can increase the reusability of the feature by slice grouping, and prevent the information loss caused by the distribution truncation caused by the feature map when grouping. Through slice grouping, we can better learn the local information of the feature map. For the selection of multiplex feature, we use convolution feature selection module, select the slice grouping features through Conv1D mapping the attention weight distribution, strengthen the useful features to weaken the useless features, and stitching the selected features into the new feature map through the way of stitching at the end.

In the field of digital image processing, the left and right neighborhood information and the upper and lower neighborhood information of pixels are very important for the entire image. For example, the traditional image feature hog feature is to use the idea to extract gradient information from the upper and lower neighbors and the left and right neighborhoods as their image features.

Assume that the parameters of the row convolution kernel are \( W_i \) Column convolution and parameters are \( W_j \). The characteristic graph is \( x_{ij} \). The calculation process is as follows

\[
\hat{x}_{\text{row}} = \sum_i W_i \ast x_{ij} \tag{3}
\]

\[
\hat{x}_{\text{column}} = \sum_j W_j \ast \hat{x}_{\text{row}} \tag{4}
\]

Adding row convolution and column convolution into the middle layer to reuse feature extraction can better enhance the fine-grained feature of the selected features, and make the extracted features more precise, thus greatly improving the accuracy of target classification. 1 * 3 and 3 * 1 row convolution and column convolution can extract the feature information of the transverse gradient and vertical gradient of the feature graph. The 24 neighborhood feature information extracted from the 5 * 5 convolution kernel and the 8 neighborhood feature information extracted from 3 * 3 convolution kernel will be more refined. However, row convolution and column convolution only apply to the relatively fine feature map that has been passed through deep convolution filtering transformation, and the effect of convolution and column convolution on the shallow feature graph or original slice is not good, compared with the square convolution kernel. The receptive field of row convolution and column convolution is relatively small, and it can not extract the global local features of coarse and fine particles. Therefore, the network depth of row convolution and column convolution added to 10 to 20 layers can better enhance the feature expression ability of the network. The calculation process is shown in Figure 12. Row convolution and column convolution are shown in Figure 13 below.
4.2. Method validation

In this paper, a 34 layer deep convolution neural network is designed as an experimental network. The ability of feature extraction for highlighting the multiplex feature selection module will be compared with the standard network on some ImageNet. The mainstream deep neural network has strong feature learning ability. Adding multiplex feature selection neural network can filter redundant features and increase the reusability of selected features.

The experiment is based on the deep convolution neural network with multiplex feature selection. After two layers of the lower sampling convolution layer, a batch normalization layer is added to ensure the stability of data distribution in the network transmission process. In the CoFS convolution network, each layer of the roll is integrated into 2 or 4 groups to convolution respectively, and the bottleneck feature is also used for linear transformation. In the training process, the convolution is about 0.5 times slower than the standard convolution speed, but the forward propagation process is about 0.5 times faster than the standard convolution. In order to improve the overall speed of the network, the parameters of the network are reduced.

The experimental network multiplexed feature selection network is trained on the TensorFlow deep learning platform. The dataset is STL-10, and the top-1 accuracy rate is calculated for the experimental results of the STL-10 dataset. SE-Net, ResNet34 and VGG-16 are selected as the contrast network of the experiment. 200 epoch is iterated to each experimental network. The experimental results are shown in Figure 14, FIG. 15, FIG. 16 and FIG. 17.
The forward propagation and backpropagation of each batch of data probably cost 224ms. Generally speaking, the multiplexing feature selection mechanism is a more effective method in deep convolution neural network. At the expense of a small amount of computation, the real weight distribution of the feature graph is learned, and then the useful features are weakened to unimportant features. It can perform feature selection very well. Multiplexing of selected features can reduce information loss and truncated distribution, and the experimental results are shown in Table 2.

| Network name | CoFS-Net (this paper) | VGG-16 | ResNet-34 | SE-Net |
|--------------|-----------------------|--------|-----------|--------|
| Training set accuracy | 95.05% | 92.88% | 94.16% | 93.79% |
| Test set accuracy | 88.52% | 85.47% | 86.94% | 84.63% |
| Training real-time rate | 224ms/step | 44ms/step | 144ms/step | 162ms/step |

From the experimental results of Table 2, CoFS-Net shows better performance than other networks, and achieves a recognition rate of 88.52% on the STL-10 dataset. From this we can see that the multiplexing feature selection mechanism has a strong ability to capture information, and the recognition rate can be improved by selecting the feature map. In order to avoid the appearance of over fitting phenomenon, the recognition rate of top-1 is improved. Dropout regularization or Earlystopping regularization techniques can be used to prevent the over fitting phenomenon in the classifier module.

In this paper, a feature extraction method based on multiplex feature selection is proposed. Aiming at the feature selection problem in deep learning, based on previous research results, convolution feature selection is proposed, and slice grouping is used to calculate attention mechanism. It increases the reusability of selected features and enhances the performance of feature extractors through the way of stitching selection features. In the middle layer, we use row convolution and column convolution to capture neighborhood information to reduce the over fitting phenomenon. Experiments show that the method proposed in this paper is slightly better than other methods.
5. Experimental setup

5.1. Experimental setup

In order to meet the needs of battlefield target recognition system, a series of orientated analysis of two demand points are carried out in this paper. The main process of the recognition module is to collect the environment target area by photoelectric sensor, extract the regional feature from the feature extraction network, and then carry out the follow-up operation of target judgment and classification through various classifiers. The specific flow chart is shown in Figure 18.

The experimental data set collected by the author is the model data set collected by the author. Among them, the positive sample training set is 2000, the test set is 200, and the size of each picture is 224 * 224, among which the target categories are divided into 4 categories, 500 for each class, and only one target in the picture. The category is human, vehicle, tank and helicopter.Rivers and other wild scenes are used to determine whether the target area contains targets. The image of the target dataset is shown in Figure 19 below.

In this paper, the feature selection network is used to extract the features of the environmental target area in the feature extraction network. The semi supervised enhanced transfer learning classifier is used to identify the target classifier in the classifier. The target classifier trains 5 identical DNN
network classifiers for fusion. In the Mac OSXchaozuo4 system, TensorFlow is deployed based on the
deep learning platform TensorFlow, and the model is compiled by Python language. The running
environment of the model is Intel Core i5, NVIDIA GTX 1060 3G GTX. The feature extractor adopts
34 layers of multiplexed feature selection network structure. The classifier uses semi supervised
enhancement of the 3 level DNN network structure after transfer learning.
In training, we first migrate the pre trained multiplex feature selection network as the feature
extraction network, then train sub classifiers of three different classification tasks according to mission
requirements, and transfer learning to the classifier through semi supervised enhanced transfer
learning. For data sets, data enhancement technology (Data Augmentation) is adopted to increase or
decrease contrast. Data enhancement is carried out by adjusting brightness and rotation translation. The
multiplication and addition calculation of the network are accelerated by cuDNN high performance
parallel computing library. The weight of the convolution kernel parameter is regularized by L2 to
reduce the occurrence of over fitting. When training the network, the SGD optimizer is used for
backpropagation computation and gradient updating. The initial value of learning rate is 0.045. With
the increase of training time, the attenuation factor is 0.01.

5.2. Analysis of results
First, the target classifier is trained. The feature extraction network trained by training is used as the
feature extractor of the experimental network in this chapter. The classifier is trained by semi
supervised and enhanced transfer learning method. The classifier network structure uses three layers of
DNN network structure, and the tail is adjusted to two output nodes to determine whether the target
area exists. For the two classification problem, In this paper, 60 percent off cross validation ROC
curves and PR curves are used as evaluation indexes to compare the two classification task classifiers
such as traditional algorithm SVM and logistic regression. The results of this experiment network are
shown in Figure 20.

Figure 20. The network experimental result
Logistic regression and support vector machine (SVM) are very good classifiers in the two
classification task. Two classifiers represent two different classification ideas. Logistic regression
follows the empirical risk minimization criterion as well as neural network, and the posterior class
probabilities are calculated by prior features. Maximizing the distance between the feature boundaries
to correctly classify the targets. The PR curves and ROC curves of 60 percent off cross validation are
also used to carry out experiments on logistic regression and support vector machines. The
experimental results of logistic regression and support vector machines are shown in figures 21 and 22.

Figure 21. Logistic regression experimental result
From figure 21, Figure 22 and figure 23, we can see that the proposed network architecture is much better than the traditional two classification algorithm logistic regression and support vector machine in judging whether the target exists or not. Under the 60 percent off cross validation test, the network structure proposed in this paper is higher than the traditional algorithm in the Precision and Recall. The area of the ROC curve is larger than that of the traditional algorithm. It can be intuitively illustrated that the method in this paper is better than the traditional method. The accuracy and recall of the sampling method after 60 percent off cross validation with logistic regression and support vector machine are shown in Table 3:

**Table 3. Experimental results of stl-10 dataset**

| Model          | 1 fold     | 2 fold     | 3 fold     | 4 fold     |
|----------------|------------|------------|------------|------------|
| Textual model  |            |            |            |            |
| Precision      | 0.57, 0.85, 1.0 | 0.54, 0.83, 1.0 | 0.8, 1.0, 1.0 | 0.71, 1.0, 1.0 |
| Recall         | 1.0, 0.46, 0.23 | 1.0, 0.15, 0.0  | 1.0, 0.25, 0.0 | 1.0, 0.42, 0.0 |
| SVM            |            |            |            |            |
| Precision      | 0.5, 0.42, 0.33 | 0.5, 0.39, 1.0 | 0.52, 0.81, 1.0 | 0.67, 0.85, 1.0 |
| Recall         | 1.0, 0.69, 0.0  | 1.0, 0.62, 0.0  | 1.0, 0.83, 0.0  | 1.0, 0.75, 0.0  |
| logistic       |            |            |            |            |
| regression     |            |            |            |            |
| Precision      | 0.57, 0.67, 1.0 | 0.54, 0.57, 0.8 | 0.49, 0.86, 1.0 | 0.75, 0.77, 1.0 |
| Recall         | 1.0, 0.61, 0.0  | 1.0, 0.76, 0.21 | 1.0, 0.75, 0.0  | 1.0, 0.41, 0.0  |

After training the target classifier, the target classifier is trained. The target classifier is essentially a multi classification task. The performance of the proposed target classifier is compared with that of the main target recognition method. The traditional target recognition method needs a classifier for each target. A classifier is used to accomplish the three tasks in the above tasks. This makes the recognition result too absolute and the target recognition rate is not high. In this paper, the relative accuracy of the target recognition method can greatly improve the recognition rate. The results of this method and VGG and SE-Net target recognition methods are shown in figures 23, 24, and 25.
This method and VGG and SE-Net methods add L2 regularization penalty term to adjust the classifier performance. The three horizontal axis is L2 regularized, the longitudinal axis is the learning rate, and the two values are log10 valued. The accuracy rate is represented by the progressive color bar on the right. From shallow to deep, the accuracy is from high to low. As shown in figures 24 and 25 above, the accuracy of SE-Net and VGG target recognition methods can reach 75.4% and 74.3% respectively in the best case of target classification tasks. This method can achieve 77.3% accuracy over other target recognition methods. The method proposed in this paper is much better than the traditional target recognition method in many classification tasks. The experimental results are shown in Table 4.

| Methods | L2 Regularization(log) | Vector(log) | Accuracy |
|---------|------------------------|-------------|----------|
| This method | 4.33 | -6.871 | 77.3% |
| SE-Net | 3.95 | -6.791 | 75.4% |
| VGG-16 | 4.07 | -6.864 | 74.3% |

6. References
[1] Chollet, François. Xception: Deep Learning with Depthwise Separable Convolutions 2016 J
[2] Xie S, Girshick R and Dollár, Piotr, et al. Aggregated Residual Transformations for Deep Neural Networks 2016 J
[3] Vapnik, V. N. An overview of statistical learning theory[J]. IEEE Transactions on Neural Networks, 1999, 10(5):988-999
[4] Wissel T, Pfeiffer T, Frysch R, et al. Hidden Markov model and support vector machine based decoding of finger movements using electrocorticography.[J]. Journal of Neural Engineering, 2013, 10(5):056020
[5] Liu W, Wen Y and Yu Z, et al. Large-Margin Softmax Loss for Convolutional Neural Networks 2016 J:507-516
[6] Schmidhuber J. Deep learning in neural networks: an overview.[J]. Neural Netw, 2015, 61:85-117
[7] X. Glorot and Y. Bengio. Understanding the difficulty of training deep feedforward neural networks. In International Conference on Artificial Intelligence and Statistics, pages 249–256, 2010
[8] Ioffe S, Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift[C]// International Conference on International Conference on Machine Learning. JMLR.org, 2015
[9] Weiss K, Khoshgoftaar T M, Wang D D. A survey of transfer learning[J]. Journal of Big Data, 2016, 3(1):9
[10] Carvalho E F, Engel P M. Convolutional Sparse Feature Descriptor for Object Recognition in CIFAR-10[C]// Brazilian Conference on Intelligent Systems. IEEE Computer Society, 2013:131-135
[11] TensorFlow:Large-Scale Machine Learning on Heterogeneous Distributed Systems[J]. 2016
[12] Hu J, Shen L and Albanie S, et al. Squeeze-and-Excitation Networks[J]. 2017Zhang T, Qi G J, Xiao B, et al. Interleaved Group Convolutions for Deep Neural Networks 2017 J

Acknowledgments
This research was supported by National Natural Science Foundation of China (61373089) and (61672360)