Traffic Sign Detection and Recognition via Transfer Learning

Liu Wei[1], Lu Runge[2], Liu Xiaolei[2]*

1. School of information engineering, chang 'an university, Xian 710064
E-mail: liu.wei@chd.edu.cn

2. Harbin Institute of Technology Shenzhen Graduate School, Shenzhen 450000, China
*Corresponding author: liuxiaolei@stmail.hittsz.edu.cn

Abstract: Automatic driving has become a extremely hot issue in recent years, and the detection of stop signs is critical for autonomous driving. Different from precious methods in which target features were extracted and then feed to SVM classifier to classify different types of traffic signs, this paper introduces a kind of transfer learning method based on the convolutional neural network(CNN). A deep convolution neural network is trained using a large data sets, and then a valid region convolutional neural network(RCNN) detection can be obtained through a small amount of traffic standard training samples. At the end of this paper, the classic GTSDB data sets and some other data of shenzhen university town are used to show the effectiveness of the transfer learning approach.

Key Words: Traffic Sign Detection and Recognition, Transfer Learning, Convolutional Neural Network

1 INTRODUCTION

Automatic driving has become a hot issue recently that not only the researchers but also the companies are devoting their energy to it. And the traffic sign detection and recognition is crucial to the autopilot system as shown in Fig.1

Fig 1. Automatic driving schematic.

Fig 2. One stop traffic sign in Shenzhen University Town.

This work is supported by National Nature Science Foundation under Grant 61170147.

Since the location of traffic signs is generally higher than the car which can be seen as Fig.2, and the color of traffic signs can not be detected by radar, the identification of traffic signs based on computer vision has important research value.

In addition, because of the special and fixed color of traffic signs, many of the results on the detection and identification of traffic signs are based on color segmentation.

When the segmentation approach is applied, a threshold need to be defined to pick out the region of interests (ROIs) among the varies color spaces.While the wildly Red-Green-Blue (RGB) color space is easy to be affected by sunlight, to solve the problem, color enhancement [1, 2] and chromatic filter are applied. Mean time, there is other method such as normalized RGB [3] is used in order to make sure the threshold. What's more, compared with RGB, Hue Saturation Intensity (HSI) [4] and Hue Saturation value (HSV) [5] are invariant to different lighting condition. And, other color spaces such as YUV [6], CIECAM97 [14] and CIELAB [27] are also used in the literature.

To sum up, the study of fast and accurate traffic sign recognition algorithm has important practical significance and application value.

2 METHOD AND EVALUATION

This paper shows how the transfer learning is applied to solve the problem of detection and recognition of stop sign. In transfer learning, first a plenty amout of images are used to train a neural network.A large-scale image training such as ImageNet [10] can be used to train a neural network. A trained neural network can solve the tasks of classification and detection well, and can then micro-train small data samples with trained networks. The advantage of using the migration learning method is that pre-trained networks have learned rich image features suitable for various images. This learning can be transferred to new tasks through fine-tuning the network. The network is fine tuned by making small adjustments to the weights so that the feature

representation learned for the original task is slightly adjusted to support the new task.

The advantage of transfer learning is that it reduces the number of images required for training and the training time. In order to illustrate these advantages, the parking sign detector is trained using a transfer learning workflow. The CNN was pre-trained using a CIFAR-10 data set with 50,000 training images. Then, this pre-trained CNN was fine-tuned to stop sign detection using only 41 training images. Without pre-training CNN, training stop sign detectors will require more images..

Note: This example requires Computer Vision System Toolbox™, Image Processing Toolbox™, Neural Network Toolbox™, and Statistics and Machine Learning Toolbox™. To train the network, GPU is suggested to use with the Parallel Computing Toolbox™.

R-CNN [12,13] is an object detection model that uses Convolutional Neural Networks (CNN) to classify image regions in images [9]. Instead of using a sliding window to classify each area, the R-CNN detector only deals with areas that may contain objects. This greatly reduces the computational cost of running CNN.

R-CNN [12,13] is an object detection model that uses Convolutional Neural Networks (CNN) to classify image regions in images [9]. Instead of using a sliding window to classify each area, the R-CNN detector only processes those areas that may contain objects. This method greatly reduces the amount of calculations generated when using CNN.

2.1 Train a CNN

Download a large scale data set which named CIFAR-10[10]. There are 50 thousands images in this dataset, which will be able to train a CNN.

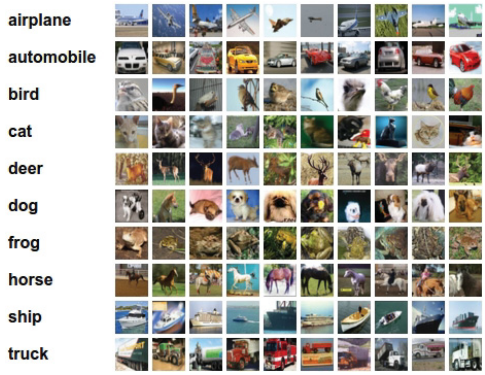

Fig 3. Examples in CIFAR-10

Neural Network is consisted of many layers, and each layer contains many neurons. Generally, a CNN is so complicated that it is hard and time - costing to design it, fortunately, the NN Toolbox is contained in MATLAB, which makes the designing and training CNN quite convenient.

In this paper, details about CNN are as follows:

The fist layer is called input layer, where the raw images are feed. As we can see, the type of input images are 32x32 RGB. Then the middle layers are consist of a series of convolution and pooling layers. In details, the activation function is Relu function. Actually, the effectiveness features can be extracted via convolution and pooling layers, which are the key points of deep learning. Also, the size of filters are defined in convolutional layers, and the parameters of filters can be updated while the CNN is on training.

In the convolutional layers, the weights of filters are defined. The random seted weights are updating while the CNN is being trained. To avoid the gradient disappear problem, Relu fuction is introduced. Meanwhile, non-linearity is added to the CNN, enabling the network to characterize non-linearities.

The convolutinal layer, RELU layer and pooling layer form the basic unit of the convolutional neural network, and multiple such units can be used to build deeper neural networks. However, the number of pooling layers should be appropriately reduced to avoid losing data. Early downsampling in the network discards image information that is useful for learning.

The final layer of CNN is usually composed of a fully connected layer and a softmax lossy layer. What's more, initialization of the first convolutional layer weight using a normally distributed random number with a standard deviation of 0.0001 helps to improve the convergence of the training.

After the network framework is defined, the network can be trained using the CIFAR-10 data set. Before training, you need to use training options to set up a network training algorithm.

Whether the training is successful should be confirmed after the network is trained. First, the rapid visualization of filter weights for the first convolutional layer can help identify any immediate issues related to training.

Fig 4. The vision of the first convolutional layer's filter

If the CNN is well trained, the weights in the first layer can form some specific shape or pattern. The random figure shows that the network still need extra training. After training, as is shown in Fig 4, the filter in the first layer have already formed specific patterns which is gained from cifar-10.

Further validation will use the data from the cifar-10 dataset to obtain a accuracy which is a indication of judging the performance of a classifier. Obviously, higher accuracy means the network is well trained. In fact, the classification accuracy of CNN can be as high as 100% in many classification tasks, while the final goal is not obtaining a high accuracy CNN. Since the final target is to gaining a object detector with a form of Region - CNN (RCNN), the efforts of fine tuning to gain a nearly 100% classification CNN is no longer necessary.

2.2 Train R-CNN Stop Sign Detector

The CNN network can identify the 10 objects well, then a stop sign detector can be obtained via transfer learning. To achieve the goal, firstly, the training dataset need to be loaded. In the training dataset, the region of interest (ROI) labels is included. Obviously, if the stop sign is what we want, then in the ROIs, only stop sign is contained.

Now that the training set contains 41 stop sign images, a R-CNN trained by such small amount of data is usually unpractical. Since the CNN has been trained on a large scale dataset which named cifar-10, then by a little fun-tune training, a reliable R-CNN detector can be obtained.

Fig 5. Training data

Finally, use the train RCNN object detector to train the R-CNN object detector. The network's input is a real image of the ground. The image contains a tagged stop sign image, using the CIFAR-10 trained CNN network and training options to complete the identification of the 10 objects. The trained R-CNN is actually a classifier. The network divides the pictures into parking signs and backgrounds.

During training, the input network weights are fine-tuned using image patches extracted from the ground truth data. The 'PositiveOverlapRange' and 'NegativeOverlapRange' parameters control which image patches are used for training. Positive training samples are those that overlap with the ground truth boxes by 0.5 to 1.0, as measured by the bounding box intersection over union metric. Negative training samples are those that overlap by 0 to 0.3. The best values for these parameters should be chosen by testing the trained detector on a validation set.

Since training R-CNN requires a lot of training resources, it is recommended to use parallel MATLAB to reduce training time. Train RCNN Object Detector automatically creates and uses a parallel pool based on the parallel preference settings. Ensure that the use of the parallel pool is enabled prior to training.

To save time, a pre-trained network is loaded from disk. It's also need to point out that if you woud like more time be saved, you'd better use GPU to trian the CNN, for the training cousumes a huge amount of calculation. After the traing, the CNN can identify the 10 kinds of object from cifar-10 well. Then, a stop sign detection can be gained through a little fien-tuning.

The R-CNN object detector returns the object bounding boxes, a detection score, and a class label for each detection. The labels are useful when detecting multiple objects, e.g. stop, yield, or speed limit signs. The scores, which range between 0 and 1, indicate the confidence in the detection and can be used to ignore low scoring detections.

3 RESULTS AND ANALYSIS

The parking marker in the test image corresponds to the maximum peak in the network activation, and it can be judged whether the R-CNN monitor can effectively identify the parking identifier. If there is more than one peak, it means that the network needs additional data for training.

Fig 6. Testing result in online data.

Fig 7. Testing result in GTSDB.

Fig 8. Testing result using the data of ShenZhen University Town.

The testing dataset is combined by 3 kind sub-dataset, which are GTSDB, online dataset and UTSZ dataset.

The testing results are shown as Table 1.

Table1. Detection results

Total signs	100
Detected signs	95
Not detected	5
False detection	1
Recall(%)	95
Precision(%)	99

It's also need to point out that the stop sign are trained in English, while in fig. 8 the Chinese stop sign can also be detected with belief score 1. Which shows that the RCNN do has a well generalization ability.

4 CONCLUSION

This paper introduces how the transfer learning is applied to solve the problem of traffic sign detection and recognition. Further details are elaborated how to train an R-CNN stop sign object detector using a CNN trained with CIFAR-10 data. To show the effectiveness of the transfer learning approach, the classic data set such as German Traffic Sign Detection Benchmark(GTSDB) are used. Furthermore, only the stop sign detector is obtained in this paper, using similar steps, other object detectors can be gained via deep learning.

REFERENCES

[1] A. Ruta, Y. Li, X. Liu, Real-time traffic sign recognition from video by class-specific discriminative features, Pattern Recognition, Vol.43, No.1, 416‑430, 2010.

[2] F. Zaklouta, B. Stanciulescu, Real-time traffic sign recognition in three stages, Robot Auton Syst, Vol.62, No.1, 16‑24, 2014.

[3] J. Greenhalgh, M. Mirmehdi, Real-time detection and recognition of road traffic signs. IEEE Trans Intell Transp Syst, Vol.13, No.4, 1498‑1506, 2012.

[4] H. Pazhoumand-dar, M. Yaghoob, A new approach in road sign recognition based on fast fractal coding. Neural Comput & Applic, Vol.22, No.3, 615‑625, 2013.

[5] C. Souani, H. Faiedh, K. Besbes, Efficient algorithm for automatic road sign recognition and its hardware implementation, J Real-Time Image Proc, Vol.9, No.1, 79‑93, 2014.

[6] J. Miura, T. Kanda, Y. Shirai, An active vision system for real-time traffic sign recognition. In: Proceedings of intelligent transportation systems, USA, 52‑57, 2000.

[7] X. W .Gao, L. Podladchikova, D.Shaposhnikov, K.Hong, N.Shevtsova, Recognition of traffic signs based on their colour and shape features extracted using human vision models. J Vis Commun Image Represent, Vol.17, No.4, 675‑685, 2006.

[8] J. F.Khan, SMA.Bhuiyan, RR.Adhami, Image segmentation and shape analysis for road-sign detection. IEEE Trans Intell Transp Syst, Vol.12, No.1, 83‑96, 2011.

[9] Girshick, Ross, et al. "Rich feature hierarchies for accurate object detection and semantic segmentation." Proceedings of the IEEE conference on computer vision and pattern recognition, 2014.

[10] Deng, Jia, et al. "Imagenet: A large-scale hierarchical image database." Computer Vision and Pattern Recognition, 2009.

[11] Krizhevsky, Alex, H. Geoffrey. "Learning multiple layers of features from tiny images.", 2009.

[12] Girshick, Ross, et al. "Rich feature hierarchies for accurate object detection and semantic segmentation." Proceedings of the IEEE conference on computer vision and pattern recognition, 2014.

[13] Girshick, Ross. "Fast r-cnn." Proceedings of the IEEE International Conference on Computer Vision, 2015.