STATE FARM DISTRACTED DRIVER DETECTION: VGG16 VS SVM

Aswinee Subramanian, Aishwarya Venkataramanan, Mi Zhou

Georgia Institute of Technology
School of Electrical and Computer Engineering
Georgia Tech Lorraine, France

ABSTRACT

One of the major causes of road accidents is the distracted driving of drivers. Based on the Kaggle competition "State Farm Distracted Driving Detection", this paper deals with the implementation of the algorithms VGG16 and SVM for detecting if a driver is distacted and classifying the the image into one of the given ten categories. VGG16 and SVM along with some of feature extraction techniques such as HOG, LBP and SURF have been implemented and their accuracy were analized.

Index Terms— HOG, LBP, SURF, SVM, VGG16

1. INTRODUCTION

As we all know, distracted driving is one of the leading causes of traffic accidents. Hence by using image technology and advanced algorithm, such accidents can be avoided greatly. However, image understanding for human is very easy but its very hard for computers. Recently, there are many approaches to classification, such as Support Vector Machine, Logistic regression, Random forest, Decision Trees and Gradient Boosting. Besides these traditional methods, deep learning has been widely accepted as an efficiently way to be used in computer image understanding. Deep CNNs are a type of artificial neural network which include convolutional filters, pooling, nonlinear activation layers, fully connected layers and the objective function loss layer [1]. At present, AlexNet, VGG, and ResNet are the models that won at the year of 2012, 2014, and 2015 respectively.

2. DATASET

We are provided with 10 classes of 640*480 pixels RGM image data for training. There are over 1000,000 images (>4GB) in the dataset, with 100 persons performing 10 different actions, which are c0: driving safely; c1: testing right; c2: talking right; c3: texting left; c4: talking left; c5: operating radio; c6: drinking; c7: reaching back; c8: hair and makeup; c9: talking to passengers. There are about 20000 images labeled in the training set and about 80000 images in the test set. The task is to label test set with probabilities for each class [2]. Some sample images are shown in Fig.1.

3. CLOUD COMPUTING ENVIRONMENT

Due to depth of VGG16, our personal computer setting is not enough. In this case, GPU 1080 Ti was used. For the operating system setting, Ubuntu 16.04.4 was used with CUDA Version 8.0.6, Python 2.7.12, Tensorflow 18.0 for running the code.

4. DEEP LEARNING NETWORK- VGG16

Transfer learning, as the name implies, is the process of transferring knowledge that has already been learned by one neural network into another one, which can be used to overcome overfitting issues and speed up the trainning process for a related work. This is also the reason we transfer the weights from ImageNet (a dataset of over 14 million images belonging to 1000 classes) to our own dataset [3]. VGG16 is a convolutional neural network model proposed by K. Simonyan and A. Zisserman from the University of Oxford[4]. The model achieved 92.7% top-5 test accuracy in ImageNet. Fig.2 shows the microarchitecture of VGG16.

A preliminary transfer learning is used to copy the learned weights from Imagenet to our own deep neural network [6]. A pre-trained VGG16 Convolutional Neural Network was used as a base layer. The last layer was removed and a dense layer with a softmax was added to output the classification. The optimization algorithm used was Adam with a small learning rate: 0.001. After trainning the model using VGG16, the behavior of the driver was predicted using this model on our test dataset. Finally the test dataset (1000 images) was labelled manually and the results were compared to get the accuracy of the model.

Thanks to Georgia Tech and Prof. Ghassan AlRegib.

c0: driving safely

c1: texting right

c2: talking right

c3: texting left

c4: talking left

c5: operating radio

c6: drinking

c7: reaching back

c8: hair and make up

c9: talking to passengers

Fig. 1. 10 different actions

5. TRAINING WITH VGG16

A preprocessing layer that takes the RGB image with pixels values in the range of 0-255 was included and the mean image values was subtracted for the entire datasets). Initially, 1000 images were used for training and got the accuracy of 55%,

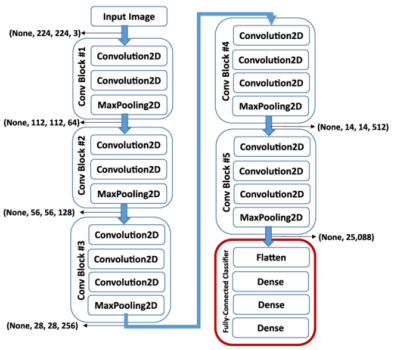

Fig. 2. The structure of VGG16

the reason may be the small size of train data. Thus 3000 samples were used to train the model. The weights trained for VGG16 on the ImageNet dataset was used for the initialization. All the layers besides the last 3 layers were frozen. The model was fine tuned using the training set and a validation set. A total of 134,301,514 parameters were used. After 4 epochs of training, the validation accuracy was 85.67%. The test dataset was used to predict the categories and the labels given by the model and the actual categories were compared. Initially a very low accuracy was obtained on test dataset, which meant that the model had a low generalization ability. When 3000 samples were used to train the model and after 4 epoches of training, the accuracy and validation accuracy reached 96.88% and 92.22% respectively, as shown in Fig.3.

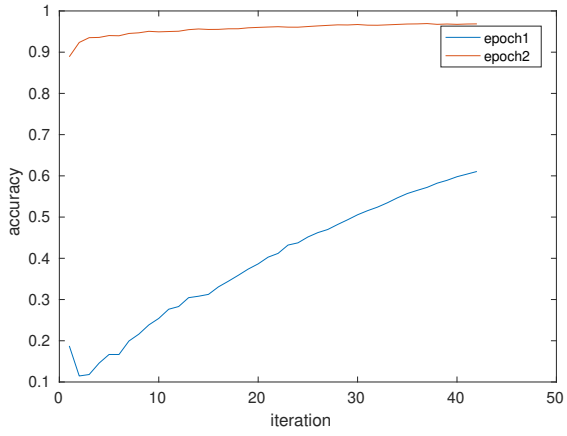

Fig. 3. The learning curve

6. TRADITIONAL FEATURE BASED METHODS + SVM

In this section traditional feature-based method combined with SVM has been used to do the classification.

6.1. HOG, LBP, SURF, SVM

The Histogram of Oriented Gradients(HOG) is a feature descriptor widely used in image processing for object detection. It counts occurences of gradient orientation in an image. A group of experiments were done to choose the best Cel lSize of extractHOGFeatures(). The cell sizes we tested were [8 8], [32 32],[64 64].

LBP, whose full name is Local Binary Patterns, is a type of visual descriptor used for classification in computer vision. LBP is very powerful combined with HOG descriptor. Function extractLBPFeatures() is used to extract LBP from images.

SURF, also known as Speeded UP Robust Features, is a patented local feature detector and descriptor. MATLAB function detectSURFFeatures() and extractFeatures() was used to do this process.

SVM is a widely-used supervised learning algorithm. In addition to performing linear classification, it can realize non-linear classification using kernel method efficiently by mapping the inputs nto high-dimensional feature spaces. In our experiment, MATLAB function fitcecoc() is used to realize the process. It can fit multiclass models for SVM.

6.2. HOG+SVM

The HOG feature of images were extracted to observe the performance, which can be realized by Matlab function extractHOGFeatures(). Two parameters of extractHOGFeatures() will influence the process of feature extracting. One of them is cell size, which means the Size of HOG cell, specified in pixels as a 2-element vector. To capture large-scale spatial information, the cell size must be increased. When the cell size is increased, small-scale details may be lost. The block size is set as default [2 2], which means the number of cells in a block, specified as a 2-element vector. Smaller block size can help suppress illumination changes of HOG features.

Table 1. The performance of using HOG+SVM.

Samples Num *traindataset*	**CellSize** *size*	**Accuracy on test dataset** *acc*
1000	[8 8]	54.9%
1000	[32 32]	56%
1000	[64 64]	50.3%
3000	[8 8]	57.1%
3000	[32 32]	58.0%
22424	[32 32]	57.4%

From Table 1, it can be found that increasing the cell size to [32 32] can improve the accuracy, but increasing it to [64 64] degrades the performance. Similarly, the number of samples in the training set also influences the accuracy. Increasing the size of the samples to 22424 does not help increase the accuracy but instead reduce it, which is very confusing for us. To make it simplifier and operational, the number of training data set is chosen as 3000. Thus, to compare with VGG16 algorithm, we choose the size of dataset as 3000, with each category containing 300 image. The cell size of extractHOGFeatures() is chosen to be [32 32].

6.3. HOG+LBP+SURF+SVM

More features were tested to see if it is beneficial to improve the accuracy. The total dimension of HOG+LBP+SURF is 2463.

Table 2. Performance using HOG+LBP+SURF+SVM.

Features	**Accuracy on test dataset**
HOG	58%
HOG+LBP	57.9%
HOG+LBP+SURF	57.5%

From Table 2, it can be seen that more features does not necessarily give a higher accuracy. It may have some degradation on the accuracy, which could be caused by the dimension of features.

7. GUI FOR CLASSIFICATION

A GUI, as Fig.4 shows, was developed which gets an input image from the user and predicts the category to which the image belongs to using a pretrained SVM+HOG+LBP+SURF network. It also displays the HOG features and the interest points in SURF. The GUI was developed using MATLAB. The detailed code can be seen in our code document, which gives a instruction how to operate it.

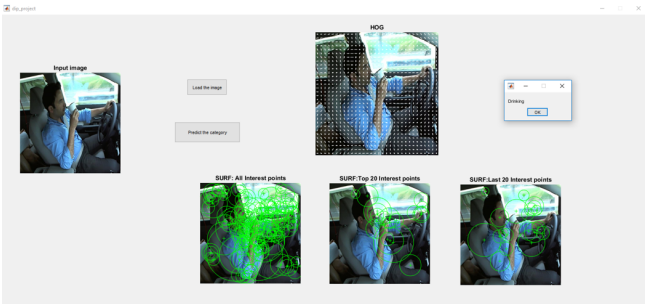

Fig. 4. The GUI

8. CONCLUSION AND IMPROVEMENT

Both VGG16 and SVM has an accuracy of 50+%, which is better than 40% and 20% in [7]. Since it is a multi-classification problem, with 10 categories, the accuracy for random guessing will be 10%, which means Machine Learning methods do work. In our experiment, the accuracy on the train dataset is around 90%, when using the trained model to predict the images that are not included in the train dataset,the accuracy is not that higher, which means the generalization of the model is poor. For the restriction of the environment, we only use 3000 images to train the model, which may be another reason for not that high predicting accuracy.

9. REFERENCES

[1] A. Krizhevsky, I. Sutskever, and G. E. Hinton, *Image net classification with deep convolutional neural networks*. in Advances in Neural Information Processing Systems 25 (F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, eds.), pp. 10971105, Curran Associates, Inc., 2012

[2] `https://www.kaggle.com/c/state-farm-distracted-`

[3] Introduction about ImageNet `http://image-net.org/`.

[4] K. Simonyan and A. Zisserman, *Very deep convolutional networks for large-scale image recognition*. Iclr, pp. 114, 2015.

[5] Singh, and Diveesh. *Using Convolution Neural Networks to perform classification on stte farm insurance driver images*. 2016.

[6] Chuanqi Tan, Fuchun Sun, Tao Kong, et al. *A Survey on Deep Transfer Learning*. The 27th International Conference on Artificial Neural Networks, 2018.

[7] Murtadha D Hssayeni, Sagar Saxena, Raymond Ptucha, and Andreas Savakis. *Distracted Driver Detection: Deep Learning vs Handcrafted Features*. Rochester Institute of Technology, Rochester, NY US. 2017, Society for Imaging Science and Technology.