Traffic Sign Classification

COMPARISON OF SYSTEMS BASED ON CONVOLUTIONAL NEURAL NETWORK AND SUPPORT VECTOR MACHINE

Viktor Šanca (*viktor.sanca@uns.ac.rs*) - Faculty of Technical Sciences, University of Novi Sad

ABSTRACT

Traffic sign recognition and classification represents one of the key aspects of autonomous vehicle development. In such a critical system it is of great importance that the classification is accurate and done in the shortest possible time. Machine vision and various machine learning algorithms have provided that the precision and speed match or surpass the human level, thus enabling rapid progress and development in the general area.

Two approaches to classification will be presented. One solution is based on ***convolutional neural network*** which exibits great potential in similar uses. Second method uses a variant of multiclass ***support vector machine*** representing a more traditional approach to classification problems.

DATA PROCESSING

Images are given in portable pixmap format (.ppm) in sizes ranging from 15x15 to 250x250 pixels. Both methods required images to be scaled to the same size. For optimized result signs were centered to the given region of interest (ROI).

Histogram of oriented gradients (HOG) was used as a feature descriptor for the SVM method and all the feature vectors have been processed before classification.

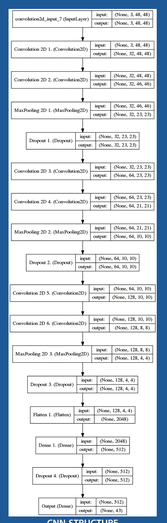
Matlab and Python have been utilized during the testing.

DATA

Dataset represents a series of more than 50 000 pictures of traffic signs of 43 different categories commonly found in europe and is part of German Traffic Sign Recognition Benchmark (**GTSRB**). This set of data is fundamental for simulating real-world conditions in *IN SILICO* experiments.

Pictures are divided into **training** and **testing** set. Training set consists of 39209 pictures in 43 classes marked from 0 to 42. Testing set contains 12630 pictures. Pictures have been collected from real world vehicular footage and have different levels of quality and noise.

CONVOLUTIONAL NEURAL NETWORK



CNN STRUCTURE

Convolutional neural network (CNN) was set up in Python using Keras and Theano as back end, running on CPU. Images were preprocessed and resized to 48x48px before given as an input to the neural network. The schematic of the implemented network is provided on the left.

Stochastic gradient descent (SGD) was used as an optimizer with the learning rate set to 0.01, decay to 10^{-6} and momentum to 0.9. To avoid overfitting images were shuffled in every epoch, while they were given in batches of 32.

Number of epochs in corelation with time needed to train the network as well as the resulting precision have been evaluated. Following number of epochs have been tested: 1, 2, 3, 5 and 20 with an average time of training of 490s per epoch.

After the training has concluded and the test has finalized the wrongly classified signs have been reviewed in order to understand the possible cause of the error. The confusion matrices have been also plotted to help with the analysis.

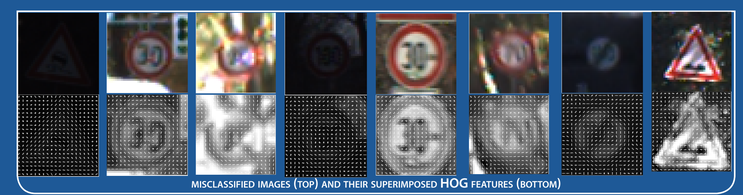
EXAMPLES OF WRONGLY CLASSIFIED SIGNS

SUPPORT VECTOR MACHINE

Support vector machine (SVM) was configured in Matlab. SVMs are by nature binary classifiers, but methods are available to extend their scope to any number of classes. This approach uses the strategy **one versus one**, by training separate classifiers for each different pair of labels, in this case a number of 43*42/2 classifiers. As this is computationally expensive it is advisable that the training dataset is optimized.

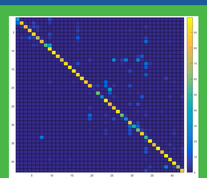
Histogram of oriented gradients (HOG) is generated for each image and such feature vector is used as an input for classification. HOG is commonly used for object detection.

The precision of the classifier is analysed through differend sizes of HOG blocks: 2x2, 4x4, 8x8, 16x16 and 20x20 pixels. The time needed for training as well as the subsequent classification results are both taken into consideration.

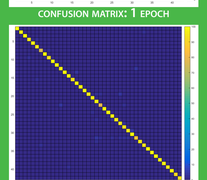


MISCLASSIFIED IMAGES (TOP) AND THEIR SUPERIMPOSED HOG FEATURES (BOTTOM)

DIFFERENT HOG BLOCK SIZES AND THEIR REPRESENTATIONS



CONFUSION MATRIX: 1 EPOCH



CONFUSION MATRIX: 20 EPOCHS

As the number of epochs increases the precision increases, with strong caution for avoiding overfitting by CNN design. The limit is reached near 20 epochs where test examples have features that are hard to distinguish and therefore classify correctly. Confusion matrices are given for the 1 and 20 epoch CNNs.

NUMBER OF EPOCHS	TIME TO TRAIN (S)	PRECISION (%)
1	490	84.6239
2	980	92.2882
3	1470	94.5210
5	2450	96.0253
20	9800	97.4743

RESULTS

Both machine learning algorithms have shown good performance in traffic sign classification. The precision of 97.4743% achieved by CNN is comparable to the human precision. It has to be noted that this classification uses real world data and some errors have to be taken into consideration, such as image noise or obstructions or low lighting conditions. It is advisable to take necessary steps to ensure that quality of images in order to achieve the highest possible accuracy of classification.

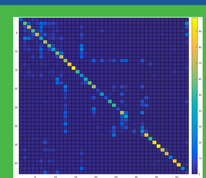
CNN has achieved the highest precision in classifying traffic signs, with a considerable drawback of long time needed for training. If taken into consideration that the neural network is itself finding features and it has a possibility to learn new classes it represents a solution that is efficient and accurate.

SVM has managed to achieve slightly lower precision for a shorter amount of time. By strictly classifying features given on the input this method reaches peak precision fast but it is unable to learn new features without restarting the whole process. Despite all the drawbacks SVM showed good performance and efficiency.

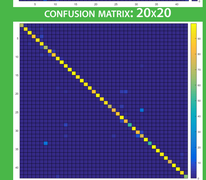
With increasing processing power and with advancements in machine learning algorithms and computer vision it is evident that vehicles will assist drivers more precisely, thus increasing the safety and security of human life.

CONCLUSION

With the decrease of the block size, thus with more details the precision continues to rise rapidly until the precision of 92% is reached. Peak precision of 92.9533% is achieved by SVM with block size of 4x4. Confusion matrices have been provided for the block sizes 20x20 and 4x4 for comparison.



CONFUSION MATRIX: 20x20



CONFUSION MATRIX: 4x4

BLOCK SIZE (PX)	TIME TO TRAIN (S)	PRECISION (%)
20x20	29.736	61.3064
16x16	26.964	72.1140
8x8	77.504	92.2011
4x4	443.967	92.9533
2x2	2420.947	92.1457

RESULTS