Development of technology for state registration signs of the vehicles recognition in conditions of low resolution on the basis of gradient bursting and convolutional neural networks

D Z Galimullin1,2, A I Denisov2, N V Denisova3, M E Sibgatullin4,5
1Kazan Innovative University, Kazan 420111, Russia
2"Veroyatnost", LLC, Kazan 420039, Russia
3Kazan State Power Engineering University, Kazan 420066, Russia
4Institute of Physics, Kazan Federal University, Kazan 420008, Russia
5Institute of Applied Research, Tatarstan Academy of Sciences, Kazan 420111, Russia

E-mail: galimullin_d.z@mail.ru

Abstract. This paper presents the technology of characters on the vehicle license plate recognition. The Viola-Jones method is used for distinguishing the area of the license plate on the basis of pre-trained Haar cascades. Convolutional neural networks analyzing the selected frames are used for recognition. Training procedure of neural networks is carried out through the created synthetic database containing characters with various types of distortion. The efficiency of neural network training was achieved due to the possibility of a sampling formation with the almost unlimited dimension, modeling of distortions which are rarely found in a real sampling due to its limitations, the possibility of forming a mixed sampling including both synthetic and real data. The necessary methods of augmentation and preprocessing for synthetic and real data are implemented, based on the resolution of the video camera. The results obtained in this work can be used for character recognition in conditions of low resolution of the video camera, distortions and lack of initial data.

The first successful applications of neural networks of deep training belong to the beginning of the 90s of the last century. The AT&T Bell Labs under the guidance of Yann LeCun developed a system of automatic recognition of handwritten figures. However, due to the complexity of convolutional neural networks training, the interest of researchers over an extended period of time has been concentrated on other approaches to machine training. Progress in the hardware and research fields has led to a renewal of interest in this method on the part of the scientific community. Particularly strong interest was witnessed after the 2010s, when such networks showed better results at recognizing images, speech, etc. At present, the systems based on the use of neural networks of in-depth training show the best results in image recognition on image test bases, such as MNIST, CIFAR-10, IMAGENET [1-3].

The paper presents the technology for the symbols of the Russian Federation vehicle license plates recognizing by the optical video flow on the basis of the gradient boosting and convolutional neural networks in the gray-scale mode and low-resolution conditions. The mechanism for synthetic training samples generating was developed and implemented, as well as a separate database of marked data relevant to the real conditions of data sampling and occurring distortions. The process of optimization, in comparison with [4], led to the network structure simplification, additional procedures inclusion,
increase of the training efficiency, reduction of the neurons number in the intermediate layers without significant loss in accuracy.

**Generation of synthetic marked out data**

The first stage in the recognition technology construction is a preliminary search for the vehicle license number, which consists in finding the area of the image containing the number. The Viola-Jones method [5-7] was used to indicate a fragment with a number plate.

It was necessary to prepare a marked out database for training the neural network. The task of preparing a high quality training sampling was one of the most important ones; besides, it was essential that the prepared data be relevant to the real data obtained during a video stream processing. To that effect, we carried out a detailed analysis of the obtained numbers database after the Haar cascades development, and detected the following distortions appearing in the frame (Fig. 1):

1. Distortions of the symbols, caused by their “blurring” due to the vehicle movement, weather conditions, considerable distance to the camera while fixing the object of interest.
2. Distortions of the background, caused by noise due to weather conditions, take-off time, camera quality, or contamination of the license plate area.
3. Distortions of symbols due to the fixation method or angle of the video camera, the form of the vehicle and method of the number fixing. Such distortions lead to geometric distortions in the symbol shape or to a change in its slope angle compared to a standard one.

Since it would be really difficult to create a real database of marked out numbers with all possible distortions included, due to the limited access to such data, it was decided to create a mechanism for generating artificial license plate numbers based on a real prototype. We should note the advantages of the synthetic training sampling. It is an opportunity to form a sampling of almost unlimited dimension, simulate those distortions that are rarely found in a real sampling, and use both synthetic and real data.

For creation of synthetic license plate numbers and simulation of distortions, we developed an algorithm [4], which served the basis for synthetic numbers generation. The example of generated images is shown in Fig. 2.

Cutting out fragments containing symbols with random distortions inserted with the help of a specially placed window was the next stage. Further, the image data was recorded into the database with the appropriate labels (Fig. 3).
Creating a training dataset

The reference [4] presents a synthesized training sample and conducted network training. The neural network showed about 99% accuracy in the test data. However, the initial accuracy did not exceed 60% when working with the real data. The carried out analysis showed that the neural network is susceptible to the quality and dimension of the input data displayed in the sliding window. If a degree of the camera resolution is less than the resolution of the images used for training, then the recognition quality is low, especially with the occurring distortions of any kind. Accordingly, it was necessary to optimize the generated synthetic data adapting them for the corresponding characteristics of the video stream images to process the real data received from the video camera. For quality improvement, it was necessary to generate samples of a larger volume with a sufficiently high diversity. It guaranteed greater stochasticity for the stable and accurate operation of the network.

In order to understand the particular distortions to be used, some of the real data was applied for validation in the synthetic-real ratio as 5 to 1. To improve the quality, synthetic data was mixed with real data from a database that was freely available on the Internet in proportions of about 5 to 1. As a result, a part of the real data comprised about 40,000 samples, including 20,000 augmented ones. The volume of the synthetic sampling comprised 250,000 samples, including 125,000 augmented ones.

The following procedures were carried out for the input data. The dimension was reduced to 16x16 pixels. This allowed speeding up the network operation and increasing the resistance to noise distortion. A unified image database was developed, which included several samplings of synthetic images obtained at different parameters mixed with the real data. The size of the final dataset made 300,000 samples.

The size of the training and validation sampling was in the ratio 4:1. Synthetic images were generated by groups with different resolutions and then were reduced to the original dimension of 16x16 pixels. In the process of synthetic data generating, a general approach to their creation was developed and the following number of transformations were used (Fig. 4):

1) Application of Gaussian noise in various proportions, application of Gaussian filters for blurring of symbols.
2) Applying distortions to basic images simulating fragments of symbols.
3) Zooming by adjusting the area of a symbol capture, resulting in a random shifts of the symbol along the X and Y axes.
4) Random turns to an angle up to 5 degrees.
5) Affine transformation with random choice of parameters in a certain range.
6) Conversion of perspective with random selection of parameters in a certain range.

The synthetic data were augmented at the next stage, using the Keras library functional and included the following procedures [8]:

1) Turns to an angle up to 5 degrees.
2) Compressing and stretching the data along the X and Y axes by 10% of the original dimensions.
3) Random shifts along the X and Y axes by 10% of the original dimensions.
4) Correlation between the neighboring colors of the pixels can adversely affect the neural network. Therefore, the ZCA Whitening procedure [9] based on the principal component method was used to eliminate correlations.

Fig. 4 Synthetic symbols augmentation example. Noise fragments are marked out with *.

Further, augmentation was carried out with the help of the same operations for the data from the real database, filmed by Russian license plate numbers and using the Keras library. The significant shortcomings of real data are due to the fact that the sizes of many symbols are fixed and occupy a certain area of the image; they have a uniform background and a fixed form, since the sampling is made under certain specific conditions. Training without augmentation here is ineffective, as the real images can be significantly distorted due to other conditions of the sampling, camera resolution, external influences. An example of symbols augmentation is shown in Fig. 5. After this procedure, the shape of the symbols, their position in the area, the color correlation on the background were significantly altered.

Fig. 5. Example of the "X" symbol after the dimension decrease. Augmentation of the real symbols.

**Recognition module optimization**

The next stage consisted in building an algorithm which allows to "collect" together all the detected symbols into a license plate, while cutting off false signals. Judging from the size of the fragment [8] the center was determined and the dimensions were set for the window. Further, when the image passed the window, the neural network determined the symbol with the maximum probability, its value and probability being recorded in the corresponding arrays (Fig. 6).
The shift was equalled to 1 pixel. Further, the probability evaluation was carried out. For each of the detected symbols, the number of its occurrences (along the X axis) and the probability for each case are shown. The Fig. 7 shows that the probabilities greater than 90% have the correct values.

Fig. 6 License plate number, symbol in the sliding window and its probability

Fig. 7. The initial image, symbols detected at scanning by the window and their probabilities.
Further, the sum of all probabilities was calculated for each registered symbol, except for the zero class marked out with * in the Fig. 6, which is the noise. All elements of the noise class were cut off. The first six, which have the maximum values, form a license plate. Together with the sum of probabilities value, they were recorded in a certain list (Fig. 8), which was indexed.

![Fig. 8 Sum of probabilities for all detected symbols](image)

Such kind of procedure allows selecting the most correct characters and numbers in the probabilistic sense. At the output after restoring the order of the coming symbols, we get the collected license plate number ['P', '7', '9', '3', 'A', 'Y'], the vector of probability sums for each symbol [3.86, 4.99, 2.9, 4.46, 4.81, 3.76], and the result of these values addition is 24.78.

The multi-window approach was proposed to reduce errors in recognition, which included the following procedure. For the purpose of the recognition authenticity increase, an additional window shift along the Y axis relative to the original position was carried out, and the characters were scanned and recognized for each window position. The shift vector was determined by the values [-6, -4, -2, 0, 2, 4, 6].

![Fig. 9 The initial image and different positions of windows. The window with optimal position is indicated separately.](image)

Figure 9 shows all possible positions of the window. We calculated the sum of the elements of the vector for each of the variants with components equal to the sums of probabilities in respect of each symbol. The maximum value indicates the optimal window position (highlighted by a rectangle), which is selected for further data processing. Later, at data processing for speeding up the analysis, we limited ourselves to a vector of shifts with the values [-6, -4, -2, 0, 2, 4, 6], since for 95% of cases this would be sufficient for correct recognition.

**Optimization of the structure and speed of the neural network**

For maximum performance, the network structure was similar to Lenet structure [2] with some changes. Its effectiveness was shown in the reference [4]. The topology of the network used is shown in Fig. 10.
Since speed was required alongside with a high recognition quality, so its structure had to be simplified as much as possible on the basis of general considerations and the work done at the previous stage. Obviously, the effective operation required the use of at least two convolutional layers, but at the same time the reduction of the number of filters without the decrease in recognition quality. It is necessary to place a decrease of dimension layer between the first and the second convolutional layers (max_pooling2d) with a minimal dimension mask, as a larger mask dimensions for images with small image dimensions shall be inefficient and lead to a strong decrease in the data dimension before the second layer. The classical ReLu-function served as an activation function. The dropout inclusion with a certain number of neuron disconnections in the Lenet-type structure made it possible to reduce the number of iterations [10].

It should be noted that this configuration, trained on synthetic data, gives an accuracy of at least 98-99% on the validation dataset, which is enough to solve our problem. In addition, the minimum number of layers provided the maximum operation speed and allowed estimating the influence of the number of convolutional neurons for the accuracy of recognition.

![Fig.10. Topology of the applied neural network](image)

The experimental estimation of operation speed and recognition accuracy for different ratios of neurons in convolutional and fully connected layers was carried out for optimizing the network structure. The used configurations, the operation speed with a single pass and multi-window mode and the recognition accuracy are shown in Table 1. Multi-window mode presents a multiple pass by a sliding window with a different height offset of the image for reduction of the recognition error. The number of shifts was equal to 7, the pitch comprised 2 pixels. The Table1 shows that the values for one pass and multi-window mode are averaged values over 40 iterations.

| Table 1. Applied topologies, their operation speed and accuracy of recognition |
|-----------------------------------------------|
| Number of the neural network structure        |
| Convolutional layer No1, number of masks      | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| Convolutional layer No2, number of masks      | 4 | 4 | 8 | 8 | 16| 16| 16| 16| 32|
| Fully connected layer, Number of neurons      | 256| 512| 512| 512| 256| 512| 512| 1024| 1024|
| Time, ms (One pass)                           | 36| 38| 41| 44| 46| 48| 52| 65| 180|
**Conclusion**

The research work has optimized the mechanism for creating the synthetic training sampling based on synthetic license plate numbers. A database of marked out license plate number symbols (according to GOST R 50577-93), relevant to the real conditions of data sampling, low camera resolution (16x16 pixels per area with a symbol) and the resulting distortions is developed. The basic configuration of the network for working with the real data is optimized in terms of speed and quality of recognition, and the training mechanism based on a synthetic learning sampling is implemented. Necessary methods of augmentation and preprocessing for synthetic and real data are implemented, based on the video camera resolution.

An algorithm for recognizing license plate numbers applying deep neural networks directly by the video stream is proposed, including a license plate number detector based on the Haar cascades, the multi-window approach for correct detection of license plate number and region symbols. The permissible license plate angle shall not exceed 10°. The recognition probability from 90% to 99% depends on the conditions of visibility and quality of license plates.

The results can be used for low-budget solutions in symbols recognition in conditions of low resolution of the camera, distortions and lack of the initial data. The approach in respect of the synthetic data use can be generalized for the other classes of objects.

The work was fulfilled under the support of the Bortnik Foundation (No. 1457GS1 / 22736).

**References**

[1] Sermanet P, Chintala S, LeCun Y 2012 *Convolutional Neural Networks Applied to House Numbers Digit Classification* International Conference on Pattern Recognition (ICPR 2012) pp.3288-3291

[2] Ian Goodfellow J, Bulatov Y, Julian I, Arnoud S, Shet V 2014 Multi-digit Number Recognition from Street View Imagery using Deep Convolutional Neural Networks (arXiv:1312.6082v4)

[3] Jaderberg M, Simonyan K, Vedaldi A, Zisserman A 2014 Reading Text in the Wild with Convolutional Neural Networks (arXiv:1412.1842v1)

[4] Galimullin D Z, Denisov A I, Illarionov S Yu 2017 Convolutional neural networks training effectiveness estimation basing on the synthetic dataset for recognition of the state registration license plate number symbols Postulate (http://e-postulat.ru/index.php/Postulat/article/view/653/676)

[5] Viola P, Jones M J 2004 Robust real-time face detection *International Journal of Computer Vision* 57(2) 137-154

[6] Wu X, Kumar V, Ross Quinlan J at all 2008 Top 10 algorithms in data mining *Knowl. Inf. Syst.* pp.1-37

[7] Chaudhari M, Sondur S, Vanjare G 2015 A review on Face Detection and study of Viola Jones method *International Journal of Computer Trends and Technology (IJCTT)* 25(1) pp.54-61

[8] https://keras.io/preprocessing/image/

[9] Kessy A, Lewin A, Strimmer K 2016 Optimal whitening and decorrelation (arXiv:1512.00809)

[10] Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R 2014 Dropout: A simple way to prevent neural networks from overfitting *Journal of Machine Learning Research* pp.1929-1958.