

## ENGINEERING AND IT

UDC 004.89

DOI 10.36074/2663-4139.11.04

# RESEARCH AND DEVELOPMENT OF TEXT RECOGNITION SYSTEM

ORCID ID: 0000-0002-5721-200X

Oleksii DENYSENKO

PE Denysenko

UKRAINE

**Abstract.** The paper discusses the technology of creating character recognition (using convolutional neural networks) systems on the image. These days, there are many approaches to solving this problem, and most of them are ineffective for images whose symbols are located on a complex background and are vulnerable to noise, affine and projection distortions. The proposed technique consists of the following stages: image pre-processing, text segmentation, and recognition by convolutional neural networks. During research was conducted a series of experiments, namely: experiment to select the most suitable method of binarization of digital images, experiment to select the most efficient convolutional neural network topology form text recognition problem. As a result of the experiments performed, this technique as applied to the recognition of car numbers demonstrates high reliability and accuracy, including in low light conditions, therefore, the developed recognition method can be recommended for commercial use. As an additional field of experiments was suggested a bunch of approaches of how to improve this technique.

**Keywords:** *machine learning, deep learning, convolutional neural networks, automatic number-plate recognition, image processing, binarization, segmentation.*

Present-day methods for recognizing text characters make it possible to solve a number of scientific as well as applied tasks, such as document recovery, publishing text on a web page, digitizing books, automating business accounting systems, determining a bankcard number. Since a number of characteristics of text data tend to change (information can be printed on images manually or using different fonts; symbols can contain digital defects or be partially displayed on images; the very images can have a complex background structure), the methods underlying software systems should provide high accuracy and speed, while remaining effective under natural conditions.

In this regard, the development of character recognition systems with a high load, which are focused on the recognition of short texts that do not have strict standards, for example, American plate numbers, is of particular relevance. Software system development is associated with a number of challenges:

1. Illumination: due to environmental influences (headlights, rain, etc.), the illumination of the source image changes.



© Денисенко О.О., 2020

© Denysenko O., 2020

<https://ojs.ukrlogos.in.ua/index.php/2663-4139><https://doi.org/10.36074/2663-4139.11.04>

2. Complex background: the background of number plates can contain drawings with complex objects that are difficult to separate from the characters in the foreground.

3. Location of the region (state): location of the identifier in the US number plates varies from state to state. This makes it difficult to generalize the methods underlying the recognition system and requires gross calculations.

4. The presence of contours, shadows, undesirable characters, etc.

The stage of the image pre-processing containing the car number plate includes image correction (removing noise from the background of the number plate, eliminating uneven distribution of brightness and the effects of focus loss) and eliminating redundant information. The pre-processing stage is no less important than all subsequent ones - the quality of image segmentation depends on its success. The method suggested in this paper uses anisotropic diffusion and equalization of the image histogram.

Since images can have a number of numerous features due to specific nature of the environment, considering one binarization method is inefficient. For better character segmentation, it is proposed to carry out binarization using a hybrid method [1]: apply five binarization methods to the source image and choose the best among them depending on the quality of the result.

#### Materials and research methods.

The global method selects a threshold value for classifying an image pixel — background or foreground [2]. The threshold value is based on the required percentage of background pixels and is calculated for a part of the image containing the necessary text information.

The Sauvola method [3] belongs to the methods of local adaptive binarization - it calculates an individual binarization threshold  $T(x, y)$  for each pixel  $(x, y)$ :

$$T(x, y) = m(x, y) \left[ 1 + k \left( \frac{\sigma(x, y)}{R - 1} \right) \right], \quad (1)$$

where  $m(x, y)$ - is the average value,

$\sigma(x, y)$  - mean square deviation at the point  $(x, y)$ ,

$R$ - maximum deviation ( $R = 128$  for shades of gray),

$k$ - the offset, which takes positive values in the range  $[0.2; 0.5]$ .

The Otsu's method [2] allows minimizing the average segmentation error that occurs when deciding whether a pixel belongs to the background or image object:

1. Calculate the sum  $P_i(k)$  for  $k = 0, 1, 2, \dots, L - 1$  by the formula

$$P(k) = \sum_{i=0}^k p_i, \quad (2)$$



where  $p_i$  - components of the normalized histogram for  $i = 0, 1, 2, \dots, L-1$ ,  
 $L$  - the maximum gray scale value.

2. Calculate the mean value  $m(k)$  for  $k = 0, 1, 2, \dots, L-1$  by the formula

$$m(k) = \sum_{i=0}^k i \times p_i. \quad (3)$$

3. Calculate the total brightness  $m_G$  by the formula

$$m_G = \sum_{i=0}^{L-1} i \times p_i. \quad (4)$$

Calculate interclass variance  $\sigma_b^2$  for  $k = 0, 1, 2, \dots, L-1$  by the formula

$$\sigma_b^2 = \frac{[m_G \times P_1(k) - m(k)]^2}{(P_1(k)[1 - P_1(k)])}. \quad (5)$$

Find the Otsu threshold as a value for which  $\sigma_b^2(k)$  is maximum.

Canny algorithm for detecting edges is often used in computer vision problems [2]:

1. Apply a Gaussian filter to the input image to remove noise.
2. Find brightness gradients by applying convolution matrices to each image pixel:

$$G_x(i, j) = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix},$$

$$G_y(i, j) = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix},$$

where  $(i, j)$  - pixel coordinates in the original image.

3. Calculate the value of the gradient  $G$  and the direction angle of the gradient vector  $\theta$  using the appropriate formulas:

$$G(i, j) = \sqrt{G_x^2(i, j) + G_y^2(i, j)}, \quad (6)$$

$$\theta(i, j) = \tan^{-1} \left( \frac{G_y(i, j)}{G_x(i, j)} \right). \quad (7)$$

4. Mark only local maxima as edges.

5. Determine the final edges by removing all the "weak" edges.



An image obtained as a result of the Canny algorithm is divided into areas that are used to calculate the local threshold. The threshold value for each block is found using the gray shade scale of the source image and neighboring ones for all boundary pixels, then the resulting binary blocks are united to create an entire binary image.

Median stacking is an image overlay technology to reduce noise, when the brightness value of each pixel is calculated by finding the median value of its brightness from a set of images. In the present paper, binarized images obtained after applying the above algorithms are suggested to be used for median stacking.

Figure 2 presents the images (1-5), which were obtained as a result of applying the five selected binarization methods to a certain source image (Fig. 1). The image of a number plate binarized by edges detection contains fewer so-called artifacts - the ratio of the pixels of the front edge to the background pixels is maximum. Thus, for the given source image, binarization based on edge detection is selected as the best among the available binarization methods.



Fig. 1. Source image of the license plate

Assuming that the region identifier is located at the top of the license plate, it is easy to search and retrieve the state name and state number using the horizontal projection method. Segmentation using projection profiles [4] takes advantage of the fact that in a binary image, the foreground pixels take the same values opposite to the background pixels. The local minima of the horizontal projection profile correspond to background pixels. The horizontal profile  $H(y)$  finds the sum of brightness of all pixels of the image  $I[x,y]$  of dimension  $M \times N$ , where  $x, y$ - are the coordinates of the pixel:

$$H(y) = \sum_x I[x,y]. \quad (8)$$

The vertical profile of  $V(x)$  projection is able to cope with symbol rotations [4]. It helps in finding the coordinates of the alleged parts of the license plate:

$$V(x) = \sum_y I[x,y]. \quad (9)$$





Fig. 2. Number plates binarized by various methods

In order to reduce memory usage and perform summation with a minimum expenditure of time, before calculating the horizontal and vertical projection values, the source image is transformed using morphological operations into a more compact version called “skeleton” [6]. The skeleton of the state plate number from Fig. 1 preserves the structure of all its objects, but removes excess pixels, as shown in Fig. 3. The extracted state name (“Texas”), which is located at the top of the number plate of the source image, is illustrated in Fig. 4.

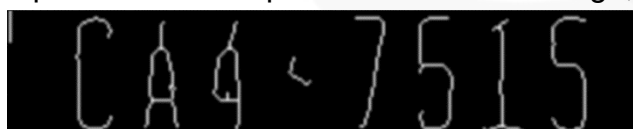


Fig. 3. Skeleton of the source image part with the state plate number



Fig. 4. State name (“Texas”) extracted from the source image

Figure 5 shows the number plate extracted using the horizontal projection method of the histogram. Character segmentation results are shown in Fig. 6.





Fig. 5. Number plate extracted from the source image



Fig. 6. Segmented number plate symbols

The convolutional neural network suggested in this paper is used as a feature extractor and classifier for recognizing the characters of the English alphabet and numbers. The number plate and the region identifier of the vehicle are recognized by individual convolution network models. The first architecture learns on binarized images and consists of three layers of convolution, pooling and normalization, as well as two fully connected layers. In the second architecture, the number of convolution layers is increased to six, and the source images are in shades of gray.

#### Research results and discussion.

Segmentation assessment was carried out on the basis of a database of 394 images. The results were compared manually by evaluating the number plate and the extracted information. 372 images were segmented satisfactorily; segmentation of 22 images failed. However, some of the images with unsuccessful segmentation were partially segmented. This gives a segmentation accuracy of number plate of 94.4%. The region identifier was extracted correctly for most images.

The first model of a convolutional neural network when learning using binary images showed a recognition accuracy of 88.97% for the region identifier - 331 recognized images out of 372 - and 15.3% for the plate number - 57 recognized images out of 372. The overall accuracy of all stages was 13% taking into account the images segmented incorrectly.

The recognition accuracy of the plate number for grayscale images was 91.12% - 339 correctly recognized images out of 372. Despite the fact that this network model has the ability to distinguish between "0" and "O", they are considered to belong to the same class due to the geometric properties of the characters. The recognition accuracy of the region identifier made 88.97%, or 331 out of 372 recognized images. The overall accuracy for all stages was 81.1% - 302 out of 394 images were successfully segmented and correctly recognized.

The competition test of license plate recognition systems, conducted by the ProSystem CCTV magazine [7], showed the results provided in the table below. Comparison of the proposed system with the closest analogues allows us to conclude that it can be used for commercial purposes with the following improvements:



1. Real-time data acquisition for processing and recognition. Real-time data includes, for example, images collected by surveillance cameras from parking lots, toll roads, etc.
2. The use of localization of the plate number.
3. Automatic execution of the stages of segmentation and localization of the identifier of the region.

The general accuracy of license plate recognition in comparison with commercial systems-analogues.

| License plate recognition system | General recognition accuracy |
|----------------------------------|------------------------------|
| "Auto-Inspector"                 | 94.19%                       |
| "Auto-Trassir"                   | 83.07%                       |
| "CVS-Auto"                       | 82.09%                       |
| "AvtoUragan-VSM"                 | 90.45%                       |
| Proposed system                  | 81.1%                        |

### Conclusion.

The use of deep learning methods in the field of recognition of text characters in images has been studied in the paper. The technology of segmentation and text recognition has been suggested in relation to the challenge of recognition of plate numbers. The most important aspects of the constructed recognition system are the introduction of a hybrid binarization method, which allowed to improve the quality of segmentation, and the use of convolutional neural networks to distinguish features and recognize license plates.

Improvement of the results can be achieved by including in the proposed method the stage of identifying a region identifier [8], as well as implementation of the stage of segmentation of a license plate using deep learning methods [9]. The implementation of all stages of the proposed system using deep learning can be time-consuming and difficult because of the limited availability of data sets. Modern graphics processors can be used as a possible solution to this problem.

### REFERENCES:

- [1] Yang, Y. & Yan, H. (2010). An adaptive logical method for binarization of degraded document images. *Pattern Recognition*, (33), 787-807.
- [2] Gonzalez, R. & Woods, R. (2018). *Digital Image Processing* (4<sup>th</sup> edition). London: Pearson Education.
- [3] Sauvola, J. & Pietikainen, M. (2000). Adaptive document image binarization. *Pattern Recognition*, 2(33), 225-236.
- [4] Kim, K.K., Kim, K, Kim, J. & Kim, H.J. (2000). Learning-based approach for license plate recognition. *Proceedings of the IEEE Signal Processing Society Workshop*, (2), 614-623.
- [5] Du, S., Ibrahim, M., Shehata, M. & Badawy, W. (2013). Automatic license plate recognition (alpr): A state-of-the-art review. *IEEE Trans. Circuits and Syst. for Video Technol*, 2(23), 311-325.
- [6] Kresch, R. & Malah, D. (1998) Skeleton-based morphological coding of binary images. *IEEE Transactions on Image Processing*, 7(10), 1387-1394.



© Денисенко О.О., 2020

© Denysenko O., 2020

<https://ojs.ukrlogos.in.ua/index.php/2663-4139>

<https://doi.org/10.36074/2663-4139.11.04>

- [7] *Recognition of license plates: "Auto-Inspector" - Test Leader*. Retrieved from [https://www.adron-perm.ru/test\\_leader/](https://www.adron-perm.ru/test_leader/).
- [8] Sermanet, P., Eigen, D., Zhang, X., Mathieu, M., Fergus, R., & LeCun, Y. (2013). Overfeat: Integrated recognition, localization and detection using convolutional networks. *arXiv preprint arXiv:1312.6229*.
- [9] Badrinarayanan, V., Kendall, A., & Cipolla, R. (2017). Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE transactions on pattern analysis and machine intelligence*, 39(12), 2481-2495. DOI: <https://doi.org/10.1109/TPAMI.2016.2644615>.
- [10] Денисенко, А.А. (2020). Исследование и разработка системы распознавания текста на изображении. *Международный журнал прикладных и фундаментальных исследований*, (5), 87-91. DOI: <https://doi.org/10.17513/mjprfi.13075>.

## ДОСЛІДЖЕННЯ ТА РОЗРОБКА СИСТЕМИ РОЗПІЗНАВАННЯ ТЕКСТУ

**ДЕНИСЕНКО Олексій Олександрович**, розробник програмного забезпечення  
ФОП Денисенко

УКРАЇНА

### Анотація.

У статті розглядається технологія створення систем розпізнавання символів (за допомогою згорткових нейронних мереж) на зображенні. В наші дні існує багато підходів до вирішення цієї проблеми, і більшість з них неефективні для зображень, символи яких розташовані на складному тлі та вразливі до перешкод шуму, спорідненості та проєкції. Запропонована методика складається з наступних етапів: попередня обробка зображення, сегментація тексту та розпізнавання за допомогою згорткових нейронних мереж. Під час досліджень було проведено ряд експериментів, а саме: експеримент з вибору найбільш підходящого методу бінаризації цифрових зображень, експеримент з вибору найбільш ефективної конвеклюальної нейромережевої топології форми розпізнавання тексту. В результаті проведених експериментів методика, що застосовується для розпізнавання номерів автомобілів, демонструє високу надійність та точність, в тому числі в умовах низької освітленості, тому розроблений метод розпізнавання може бути рекомендований для комерційного використання. Додаткового було запропоновано кілька підходів щодо вдосконалення цієї методики.

**Ключові слова:** машинне навчання, глибоке навчання, згорткові нейронні мережі, автоматичне розпізнавання номерів, обробка зображень, бінаризація, сегментація.

