Intelligent information systems

UDC 004.032.26, 528.854

DOI: https://doi.org/10.20998/2522-9052.2023.1.09

Andrii Podorozhniak, Natalia Liubchenko, Maksym Sobol, Daniil Onishchenko

National Technical University “Kharkiv Polytechnic Institute”, Kharkiv, Ukraine

USAGE OF MASK R-CNN FOR AUTOMATIC LICENSE PLATE RECOGNITION

Abstract. The subject of study is the creation process of an artificial intelligence system for automatic license plate detection. The goal is to achieve high license plate recognition accuracy on large camera angles with character extraction. The tasks are to study existing license plate recognition techniques and to create an artificial intelligence system that works on big shooting camera angles with the help of modern machine learning solution – deep learning. As part of the research, both hardware and software-based solutions were studied and developed. For testing purposes, different datasets and competing systems were used. Main research methods are experiment, literature analysis and case study for hardware systems. As a result of analysis of modern methods, Mask R-CNN algorithm was chosen due to high accuracy. Conclusions. Problem statement was declared; solution methods were listed and characterized; main algorithm was chosen and mathematical background was presented. As part of the development procedure, accurate automatic license plate system was presented and implemented in different hardware environments. Comparison of the network with existing competitive systems was made. Different object detection characteristics, such as Recall, Precision and F1-Score, were calculated. The acquired results show that developed system on Mask R-CNN algorithm process images with high accuracy on large camera shooting angles.

Keywords: license plate detection; object detection; machine learning; deep learning; convolutional neural network; region based convolutional neural network; Mask R-CNN.

Introduction

Intellectual computer systems are everywhere in our lives: from online shopping to space exploration. One example of many such systems are automatic license plate recognition systems (ALPR). ALPR allows making our life easier, more comfortable and most important safer. According to Ministry of Internal Affairs of Ukraine [1] in 2022, 237 smart cameras are installed (Fig. 1). Main core of these systems is ALPR module, which makes possible extraction of license plate number. These systems can be used not only as part of speeding detection systems but also by automatic parking, access control, law enforcement or accident detection systems. Market of ALPR systems was assessed to be 774 million (USD) in 2019 with rise to 1885 million (USD) in 2027 [2]. Newer research shows, that market price was 1050 million (USD) in 2021 and will be raised to 1931 million (USD) in 2028 [3]. Both researches show that market value of ALPR technologies as much as their demand will grow.

Problem statement

In terms of Computer Vision, this problem belongs to object detection problems. In nutshell, ALPR task can be declared as procedure of finding coordinates of license plates on image. For creation of the system, many more factors should be taken into account.

Firstly, it is important to understand, how images are acquired [4, 5]. Because of different camera angles sometimes license plate recognition simply impossible (Fig. 2). Another common restriction for these systems is weather. Snowy or rainy day images sometimes harder to understand than night images.

Secondly, stages of image processing must be identified [6] – whatever to use preprocessing or not, how exactly processing of the image will look like, what about post-processing and so on. All these questions are part of main method selection. For this purpose, different
approaches can be applied – RGB to grayscale transition or histogram normalization for preprocessing, corner detection methods for image processing, natural language processing (NLP) in terms of post-processing for symbols extraction. Because of low accuracy of these methods [7 – 9], artificial neural network was chosen [10, 11].

Standard ANN have no practical results for image analysis, that’s why convolutional neural network is used. CNN is a class of deep learning techniques for Computer Vision problems. In terms of object detection, R-CNN is a good class of convolutional neural networks. More specifically, Mask R-CNN was applied.

**Mask R-CNN**

Mask R-CNN is based on region proposal networks and improvement over Faster R-CNN [12] for semantic segmentation. In presented work, Mask R-CNN used for license plate detection purpose.

Mask R-CNN consist of two parts [13]. In the beginning, convolutional neural network performs feature map extraction. CNN can be chosen freely (VGG16, VGG19, ResNet50, ResNet101, InceptionNet, DenseNet or EfficientNet), in our case, ResNet101 was used. Simultaneously, regions are proposed. These regions contain the objects in feature maps.

Second part of the network is bounding boxes prediction and object classification module. The module works once pro region obtained in stage 1.

As mentioned before, Mask R-CNN is improvement of Faster R-CNN (Fig. 3) for instance segmentation. As the result, created network predicts not only license plate position, but also mask.

**Fig. 3. Faster R-CNN general architecture**

Because of the combined architecture of Mask R-CNN (Fig. 4), accuracy characteristics of the network are much better, than previous networks [13]. Important to notice, that main advantage of chosen system is, in the same time, main problem – complex loss function of the network.

Loss function of the network consists of three parts: classification loss, localization loss and mask loss and can be described like:

\[ L(u, p, t^p, v^u, p) =
L_{cls}(p, u) + \lambda L_{loc}(t^p, v^u) + L_{mask}(m, u, p). \]

To fully grasp this function behavior, it is important to look at all part losses. Each sub-loss is not so hard to understand separately:

\[ L_{cls} = -u \cdot \log(p) - (1 - u) \cdot \log(1 - p); \]

\[ L_{loc}(t^p, v^u) = \sum_{i \in \{x, y, w, h\}} L_{smooth}(t^p_i - v^u_i); \]

\[ L_{mask}(m, u, p) = -\frac{1}{m^2} \sum_{i=1}^m \sum_{j=1}^m \left[ u_{ij} \cdot \log(p_{ij}) + (1 - u_{ij}) \cdot \log(1 - p_{ij}) \right], \]

where:

- \( u \) – class of the object;
- \( u_{ij} \) – class of the object in pixel \( i, j \);
- \( p \) – predicted class of the object;
- \( p_{ij} \) – predicted class of the object in pixel \( i, j \);
- \( v^u \) – real bounding box;
- \( t^p \) – predicted bounding box;
- \( \lambda \) – coefficient of the localization loss function weight (originally is 1);
- \( m \times m \) – size of the mask;

\[ L_{smooth}(x) = \begin{cases} 
0.5 \cdot x^2, & |x| < 1; \\
|x| - 0.5, & |x| \geq 1.
\end{cases} \]

**Metrics evaluation**

To test accuracy of developed ALPR system [14] standard Accuracy metric was used:

\[ Accuracy = \frac{TP + TN}{TP + TN + FP + FN}, \]

where:

- \( TP \) – number of true positives (number plate is presented and detected);
- \( TN \) – true negatives (number plate isn’t presented and isn’t detected, not applicable to the problem, equals zero);
- \( FP \) – false positives (number plate wasn’t presented, but was detected);
- \( FN \) – false negatives (number plate was presented but not detected).

Accuracy comparison developed system to NumeroffNet [15] and SeeAuto (FF-Group) [16] in a wide range of camera shooting angles (Fig. 5) was made [14, 17].
Results show, that developed system works better, than competitive solutions, especially in terms of large camera shooting angles.

Another way to test presented system is to use Precision, Recall and F1-Score metrics (Table 1).

Table 1 – Precision, recall and F1 score of the model

<table>
<thead>
<tr>
<th>Metrics</th>
<th>Precision</th>
<th>Recall</th>
<th>F1-Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Autoria Number Plate Dataset</td>
<td>0.96</td>
<td>0.95</td>
<td>0.95</td>
</tr>
<tr>
<td>Medialab LPR dataset</td>
<td>0.96</td>
<td>0.93</td>
<td>0.94</td>
</tr>
</tbody>
</table>

For this purpose, two datasets were chosen: Medialab LPR dataset [18] and Autoria Number Plate Dataset [15].

\[
\text{Precision} = \frac{TP}{TP + FP}; \\
\text{Recall} = \frac{TP}{TP + FN}; \\
F1 = 2 \cdot \frac{\text{Recall} \cdot \text{Precision}}{\text{Recall} + \text{Precision}}
\]

Precision and recall are more useful than accuracy if some type of errors are more valuable than another. For instance, if it is important to minimize false detection of number plates – precision is the main metric.

Unlike accuracy, F1 score is balancing precision and recall, and not simply saying, how accurate the system is. F1 is based on “wrong” answers and calculated through error; on the other hand, accuracy is based on “correct” answers.

**Hardware implementation**

Strictly speaking, there are two possible implementation types for ALPR systems – software and hardware [19]. A software implementation is a computer program, which receives an image as input and generates an array of coordinates of license plate positions. Hardware implementation is, for example, a smart speeding detection camera. Software based systems are more general-purpose and don’t depend on environment, where they installed. Hardware based solutions are more powerful and can perform real-time on-the-spot detection, but they are also more expensive.

Another possible solution is combined implementation of the system [20]. It means, that camera transfers to server preprocessed image and main computational algorithm is executed on a powerful mainframe.

This restriction is made, due to lack of computation capabilities of hardware platforms. As a platform for the project, Raspberry Pi was chosen, because of availability and community support. Most of the DIY projects are based on OpenCV processing, such as RGB to grayscale transformation and different edge detectors [21, 22]. All these methods don’t come even close to accuracy that can be achieved, with the help of Mask R-CNN. That’s why Raspberry Pi implementation of the proposed system was created (Fig. 6).

As expected, first results weren’t great. One image could be processed from 10 to 20 seconds (compared to average 4 seconds with Intel I7-9750H CPU). The main framework is TensorFlow, which makes deep learning possible, but it is very large for chosen platform. That’s why TensorFlow Lite was used. Because visualization part (Fig. 7) was made with the help of OpenCV, OpenCV Lite was selected. After all optimization steps, average processing time decreased to 7 seconds.
Although no real-time processing is possible with current platform or without model optimization, realization of the solution can be implemented with different processor and build in ALPR drone (Fig. 8) [23, 24].

Conclusions

The aim of the work is analyze of theoretical background and practical problems of automatic license plate recognition systems.

The article defines the problems of ALPR system creation and presents possible solution. Different approaches were analyzed and main detection algorithm was selected. Mathematical background of Mask R-CNN was presented and clarified. After that, developed system was tested in comparison to different rival systems.

Important to notice, that processing time with GPU (GeForce GTX 1060TI) takes only 300 ms, it means, that the system is great solution for server-based recognition.

Results show that created system works better than compared network on big camera shooting angles. Different accuracy metrics were calculated to show the system capabilities.

Furthermore, created system was implemented in different hardware environment. Results show that developed network can be applied in server-based solutions and doesn’t allow real-time analysis.

As a result, an effective system for automatic license plate recognition was developed, presented and tested. Accuracy more than 95% was achieved for viewing angles up to 60 degrees. High recall and precision metrics show balanced identification possibilities of the network.

REFERENCES

Використання Mask R-CNN для автоматичного розпізнавання номерних знаків

А. О. Подорожняк, Н. Ю. Любченко, М. О. Соболь, Д. П. Онищенко

Анотація. Предметом дослідження є процес створення системи штучного інтелекту для автоматичного визначення номерних знаків. Мета полягає в тому, щоб досягти високої точності розпізнавання номерних знаків під великими кутами камери з виділенням символів. Завдання полягає у дослідженні існуючих технологій розпізнавання номерних знаків і створення системи штучного інтелекту, яка працює на великому кількості кутів зйомки за допомогою сучасного широкого діапазону відео. У роботі проведено аналіз і вивчення разних методик розпізнавання, які використовуються у сучасних системах розпізнавання. У результаті аналізу, висновку з точки зору точності та швидкості обробки, було вибрано алгоритм Mask R-CNN, який завдяки своєму принциповому прийому вибірки, забезпечує ідеальну точність. У рамках дослідження було вивчено та розроблено апарати, які використовуються для обробки зображень. Як результат, система має можливість розпізнати різні номерні знаки з низької кількості кутів камеры. Надійшла (received) 08.12.2022

Прийнята до друку (accepted for publication) 15.02.2023

Daniil Onishchenko – кандидат технічних наук, доцент кафедри комп’ютерної інженерії та програмування, Національний технічний університет "Харківський політехнічний інститут", Харків, Україна;
Andrіїн Podorожняк – Кандидат технічних наук, доцент кафедри комп’ютерної інженерії та програмування, Національний технічний університет "Харківський політехнічний інститут", Харків, Україна;
Nataliia Liubchenko – студент кафедри інформатики та інтелектуальної власності, Національний технічний університет "Харківський політехнічний інститут", Харків, Україна;
Maksym Sobol – Кандидат технічних наук, доцент кафедри комп’ютерної інженерії та програмування, Національний технічний університет "Харківський політехнічний інститут", Харків, Україна;
Daniil Onishchenko – Кандидат технічних наук, доцент кафедри інформатики та інтелектуальної власності, Національний технічний університет "Харківський політехнічний інститут", Харків, Україна;
Nataliia Liubchenko – Кандидат технічних наук, доцент кафедри інформатики та інтелектуальної власності, Національний технічний університет "Харківський політехнічний інститут", Харків, Україна;