UDC 004.93
IRSTI 28.23.15

https://doi.org/10.55452/1998-6688-2024-21-3-66-77

[1]*Vykhodtseva V.A.,
Master student, ORCID ID 0009-0007-1897-0768,
e-mail: vykhodtseva.va@gmail.com
[1]Popova G.V.,
PhD, Associate Professor, ORCID ID 0000-0002-6935-1066,
e-mail: gpopova@edu.ektu.kz

[1]Kazakh-American Free University, 070000, Ust-Kamenogorsk, Kazakhstan

APPLICATION OF MACHINE LEARNING TECHNIQUES TO INCREASE THE LEVEL OF ACCURACY OF OPTICAL CHARACTER RECOGNITION RESULTS

Abstract

One of the most pervasive processes of modernity is undoubtedly digitalization, which has encompassed all key spheres of human life. The development of information technology has contributed to large-scale changes not only in the everyday aspect of life, but also more globally, automating complex business processes in the field of entrepreneurship, economics, and healthcare. The transition to digital data and documentation has provided greater accessibility to necessary information and has also enhanced the efficiency of its analysis and processing. Due to this fact, optical character recognition (OCR) technology has gained significant importance, enabling the identification and extraction of textual data from images. OCR systems play a pivotal role in the digital transformation of society as they eliminate the need for manual handling of textual information in images and are applicable in automating the majority of business processes associated with paper-based data processing, such as gathering statistical data from paper forms, reflecting paper documents in electronic document management systems, converting textual information into audio files, and so on. This paper is dedicated to describing optical character recognition technology, as well as providing an overview of machine learning techniques that are actively used in the context of its modern implementation, in order to enhance the quality of the obtained results. In addition, the paper presents the principles of operation of the described approaches, their capabilities, as well as some limitations that may be encountered when using them in various scenarios.

Key words: optical character recognition, features, machine learning, deep learning, convolutional neural network, recurrent neural network.

Introduction

In the digital age, where key business operations rely on processing large volumes of information, the ability to quickly and accurately extract valuable information from text data is of paramount importance. This is why many large enterprises have become interested in implementing optical character recognition (OCR) technology. This is due to the proliferation of unstructured text data, such as invoices, receipts, contract forms, and therefore it poses a challenge for such organizations seeking efficient, streamlined business processes and high competitiveness. Optical character recognition is a technology that enables the recognition and conversion of data located in images into machine-readable text. The data obtained in this way can be analyzed or processed using appropriate software. Thus, there is no need for manual data entry from images for their subsequent use as a text format [1]. As this mechanism began to be more actively studied, the possibility of integrating optical character recognition and deep learning techniques was discovered, which revolutionized the accuracy, efficiency and versatility of text recognition systems. Using the power of neural networks and data-driven learning, optical character recognition systems have overcome long-standing limitations and

are now able to identify text with unprecedented accuracy and efficiency. This combination is capable of recognizing handwritten text in images and analyzing printed documents with complex fonts. The use of deep learning models has opened up unprecedented possibilities, taking OCR technology to new levels of performance and applicability.

Literature Review

In recent years, issues related to optical character recognition technology have become increasingly prevalent both within the IT and research communities [1–7]. Early research in optical character recognition mainly relied on classical algorithms to perform character recognition tasks. However, the advent of deep learning has revolutionized the field by introducing a data-driven approach that learns complex patterns and representations directly from raw input data. However, the advent of machine learning, in particular deep learning, has revolutionized the field by introducing an approach based on automatically learning complex patterns and representations directly from raw input data. It is issues related to the implementation of the integration of optical character recognition and deep learning techniques that have become the main focus of many studies. Papers such as [8, 10, 12] have demonstrated the effectiveness of neural networks in developing systems capable of performing character recognition on images with complex structures and specific characteristics. One of the most important issues is the recognition of handwritten text, which was highlighted in studies [13, 14, 16, 17] that explored the solutions of the problem using recurrent and convolutional neural networks. Studies [19, 20] reveal the features of a hybrid architecture consisting of these two neural networks and its application in complex tasks. Papers [21–25] are dedicated to the Tesseract system, which embodies all the capabilities inherent in a modern optical character recognition system, including learning from training data using neural networks. They demonstrate the practical results of using this mechanism.

Main provision

Optical character recognition technology includes the following stages of image processing:

1. Image preprocessing. This stage involves operations to enhance the quality and legibility of the image: it aligns the image, reduces background noise, and increases the contrast of relevant areas. At this step, binarization is also performed, which is converting the image into a format consisting of 0 and 1 and visually rendering the image in black and white [2];
2. Segmentation. This is the step at which the mechanism analyzes the provided image and identifies areas with text, lines, distances between words, and other attributes [3];
3. Character recognition. At this stage, the technology performs a series of procedures to identify each character and convert it into text. There are two algorithms that are possible in this case: pattern recognition and feature detection;
4. Post-processing and correction. If errors or distortions are detected, the system corrects them and takes additional steps to improve the quality of the recognized text. At this stage, individual characters are combined into words, and text formatting is performed to ensure its structured presentation [4].

The sequence of steps is more clearly illustrated in Figure 1.

Concerning the above-mentioned classic OCR technology algorithms, in pattern recognition, a set of pre-prepared image samples or character templates stored in memory is utilized to compare against the corresponding character in the image [5]. This approach operates quite successfully under conditions in which the image possesses high quality without significant distortions, and the text contained within it is clear for recognition. The feature detection method, in turn, is based on the characteristics of each character, such as how many lines it contains, whether these lines intersect, the proportions of the character, and so on [6]. The characters are subsequently classified according

to the identified features, and character recognition occurs based on this classification. In case the technology supports multiple languages, additional classification based on fonts and alphabets is carried out.

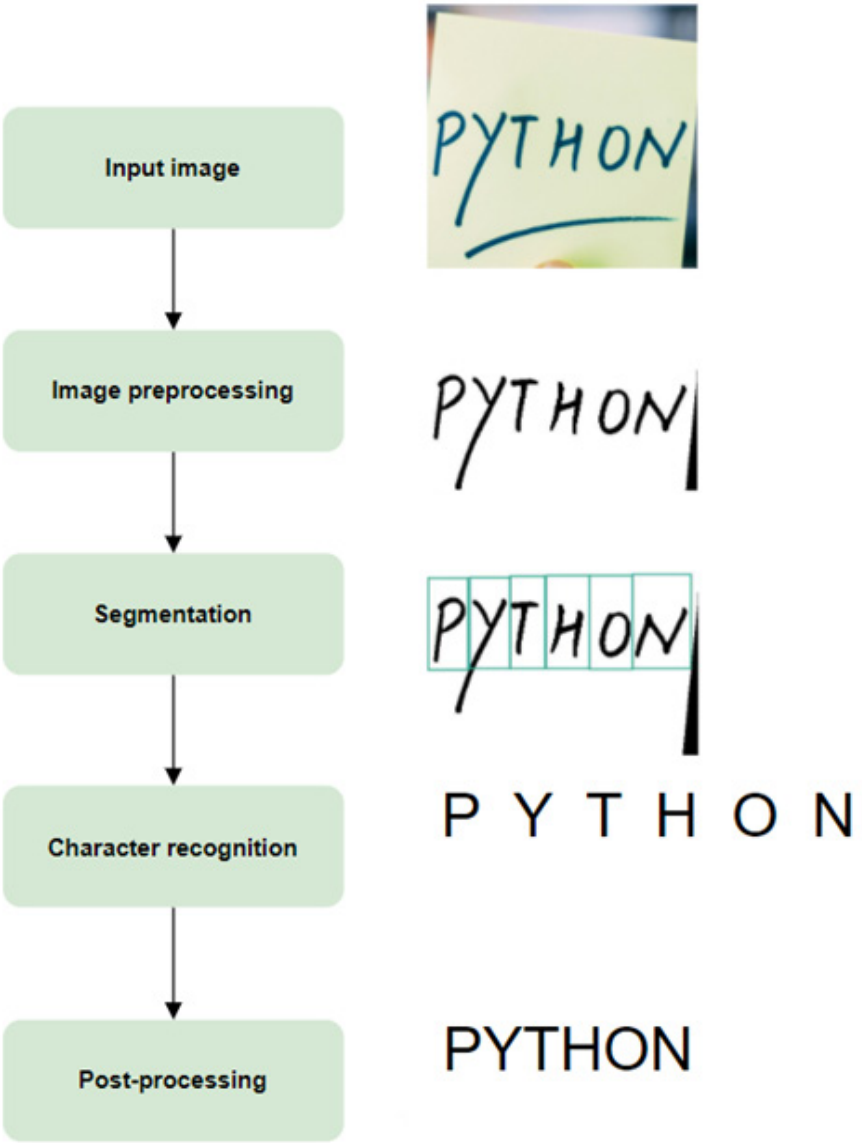

Figure 1 – Stages of operation of an OCR system

In Figure 2, this algorithm execution process is presented in the form of a generalized block diagram.

Despite the fact that these algorithms are capable of detecting and converting text in images, the quality of their output can vary significantly depending on the characteristics of the input image: the noise level, font, language of the text, and so on [7]. For this reason, optical character recognition technology is increasingly being combined with machine learning techniques, since such a combination has demonstrated the most accurate results in text recognition in images [8].

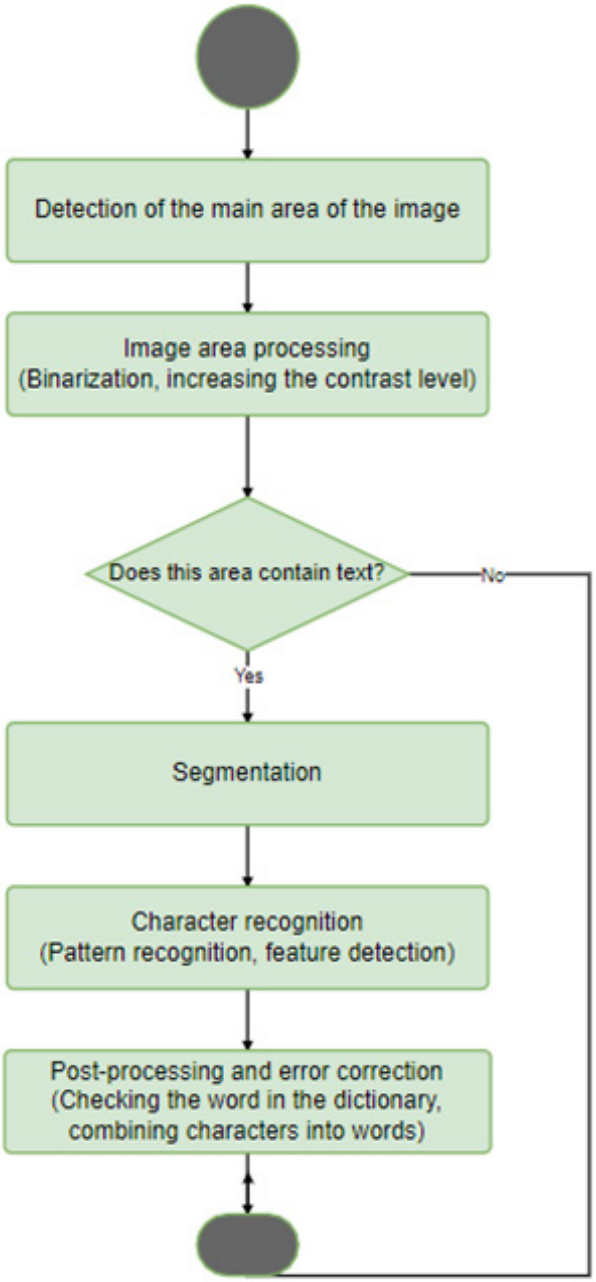

Figure 2 – Algorithm of an OCR system

It should be noted that machine learning is a subfield of artificial intelligence that enables a system to learn from training data — the better the model is trained, the higher the accuracy of character recognition, even on complex images [9]. When using machine learning techniques in an OCR algorithm, a large amount of labeled data is required to train the models, since the appearance of images belonging to the same category can be presented in different ways, for instance, the arrangement of fields with the same text may differ across different image samples [10]. Therefore, images with non-standard elements or complex structures may be an obstacle to the effective decision-making of machine learning models. A general scheme of how machine learning works is presented in Figure 3.

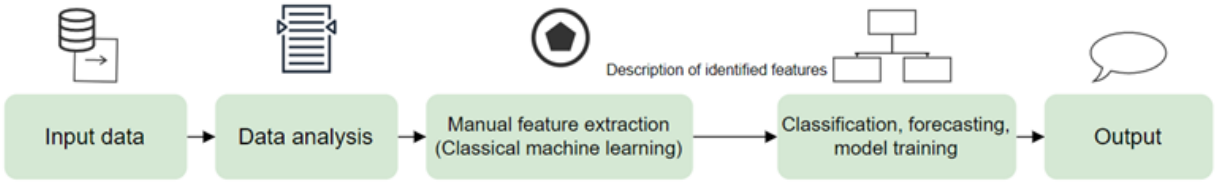

Figure 3 – Stages of classical machine learning

In order to make optical character recognition algorithms more flexible, deep learning has become even more widespread. It is a subfield of machine learning that utilizes multi-layered neural networks to automatically extract complex hierarchical features from data, rather than just using samples as training data [11]. Neural networks are a collection of many interconnected nodes that interact with each other during the process of data processing. Each node in such a neural network is responsible for solving a specific task, and after its completion, the processed data is transferred to the next node.

Materials and methods

Deep learning demonstrates high accuracy and robustness to various image characteristics, such as noise, blur, scale variation, and distortion [12]. As for limitations of this approach, it should be noted that such models operate only with a large amount of training data and require significant computational power for complex mathematical calculations. Damaged or incomplete data can therefore affect the accuracy and quality of the obtained result. The general sequence of steps in deep learning is schematically presented in Figure 4.

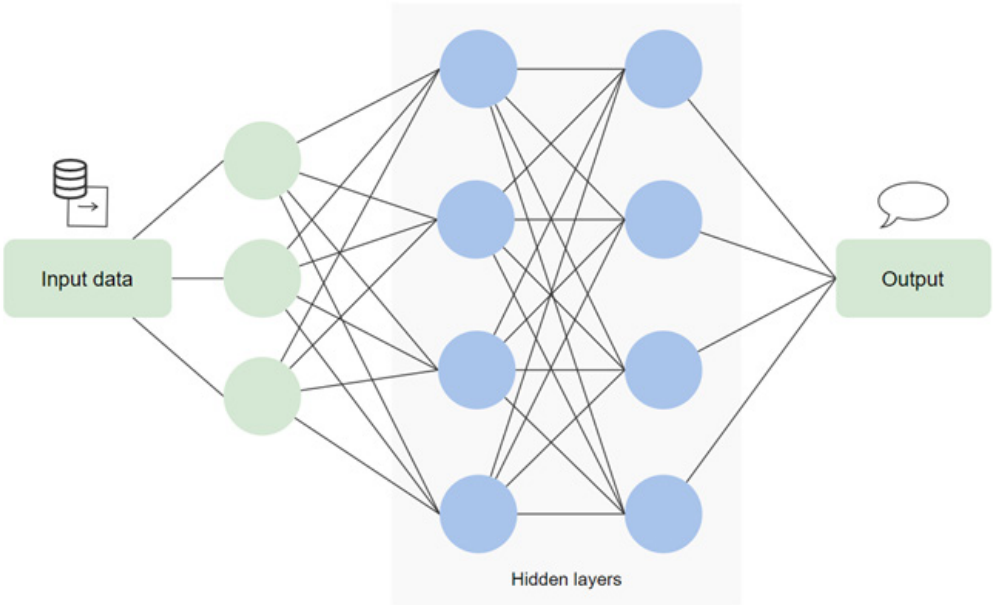

Figure 4 – Stages of deep learning

In the context of OCR technology, as a rule, two types of neural networks are used for text recognition in images: convolutional neural networks (CNN) and recurrent neural networks (RNN) [13]. Their usage in solving this problem ensures the highest level of accuracy in character recognition, even on very difficult-to-read images.

Recurrent neural networks are used to process sequences efficiently. They process text in an image as a sequence of characters, capturing contextual dependencies, since their architecture involves loops that allow previous states to be remembered when processing new data. Information about previous data is transmitted using hidden states, which is the internal state of the network at the current time step. At each such time step, the input data and the current hidden state are used to compute a new hidden state, and this new hidden state is passed on to the next time step and used again to analyze the next element of the sequence [14]. When dealing with different languages, such a neural network can be used in combination with language models to ensure text recognition in images, taking into account the linguistic context. Recurrent neural networks are often used in tasks related to text analysis when translating into another language, time series, sequence generation, and audio signal analysis.

The update of the hidden state h_t at time step t can be represented by the following mathematical formula:

$$h_t = \sigma(W_{hx}x_t + W_{hh}h_{t-1} + b_h) \quad (1)$$

Where x_t is the input vector at the current time step, h_{t-1} is the previous hidden state, W_{hx} и W_{hh} are the weight matrices for the input and hidden states, respectively, b_h is the displacement vector, σ – activation function.

A general and detailed diagram of a recurrent neural network is presented in Figure 5.

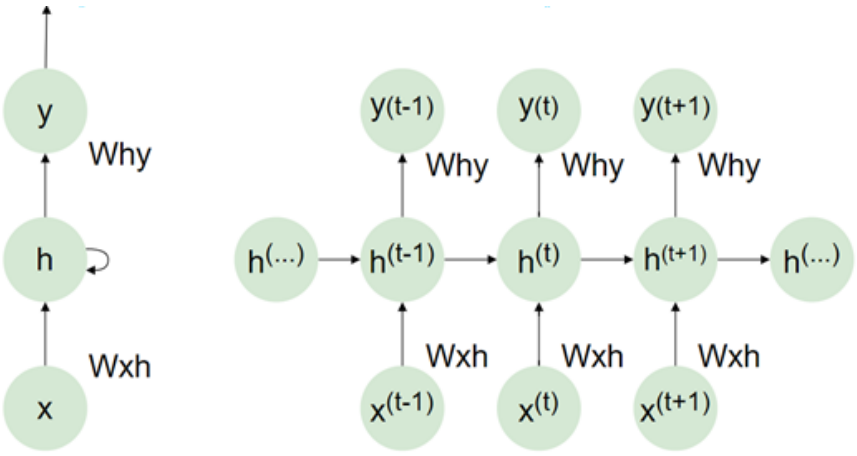

Figure 5 – Recurrent neural network

In the context of optical character recognition, a recurrent neural network takes a sequence of input vectors, which are the characters of the text in an image, and then updates its internal state at each step [15].

Convolutional neural networks are also used in OCR systems to effectively extract characteristic features of text in an image, since they are adapted to work with data that has many coordinates [16]. Their architecture consists of:

1. A convolutional layer to extract significant features, for example, character outline, texture, character shapes;
2. Pooling layer to reduce the dimensionality of the output data;
3. Flatten layer as an activation function to form more complex functions at the output of the layer;
4. A fully connected layer is usually the final layer in which each neuron is connected to all neurons of the previous layer [17].

As mentioned above, in the convolutional layer, a neuron is connected to some local area of the previous layer, and in a fully connected layer, a connection is made to all neurons. The convolution operation in CNN can be represented mathematically as follows:

$$(I * K)(i, j) = \sum_m \sum_n I(i + m, j + n) K(m, n) \qquad (2)$$

Where I is the input image, K is the convolution kernel (filter), (I * K)(i, j) is the result of convolution at position (i, j) [18].

Within the task of text recognition in an image, this operation allows for the calculation of the weighted sum of pixel values in the input image, overlaid with a convolution kernel. The result of convolution is a feature map containing information about spatial patterns in the image.

In addition to text recognition in images, convolutional neural networks are often used in tasks such as image classification and recognition of objects or faces within them.

The general scheme of operation of a convolutional neural network is presented in Figure 6.

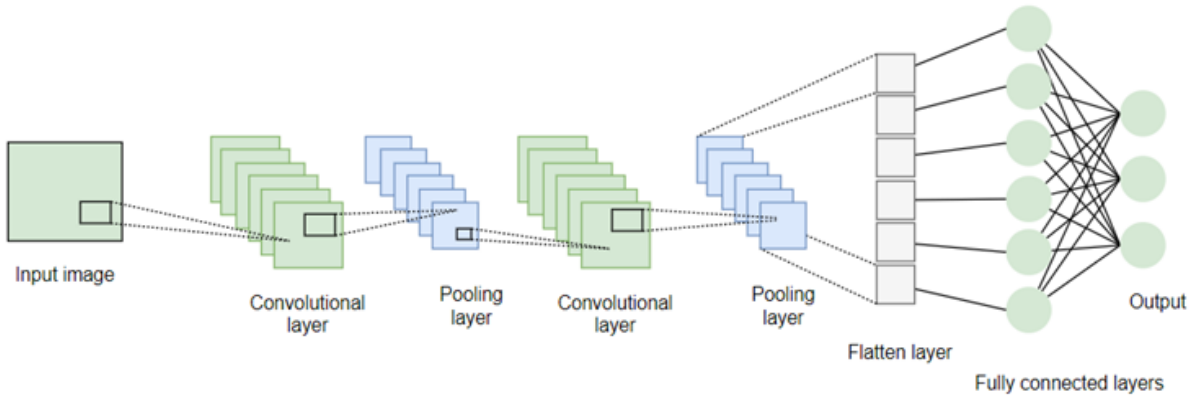

Figure 6 – Convolutional neural network

In addition, a convolutional neural network can work in the same architecture as a recurrent neural network. The combination of these two architectures, called convolutional recurrent neural network (CRNN), gives better results when processing images containing text. It consists of three components: a convolutional layer, followed by a recurrent layer, and a transcription layer [19]. Based on the previously described properties of these neural networks, the CNN layer is used to extract spatial features from images, then the obtained results are transferred to the RNN layer to analyze the sequence of characters taking into account previous contexts. This allows for the efficient formation of complete words and sentences from each recognized character based on this relationship [20]. In addition, these layers of the architecture are trained synchronously, which allows the model to effectively work with both local and global image contexts. This architecture is quite complex and demanding on computing resources, but it is more universal for a wide range of tasks for recognizing characters in an image. Convolutional recurrent neural network is often used in handwritten text analysis tasks. A general diagram of the functioning of the CRNN architecture in the context of OCR is presented in Figure 7.

Based on the above, it should be noted that the choice of a particular technique depends on the type of input data intended to be used. Complex fonts, low image quality, characteristics of alphabets, and unpredictable placement of the same elements on different images significantly affect the results of OCR system operation. In cases where basic conversion of text in an image into a machine-readable format is required, or when only high-quality and easily readable images are used, classic algorithms are also suitable, since the results in this case are quite predictable and the conversion process is more likely to not require complex processing algorithms, which, for example, offer machine learning

techniques. However, the application of machine learning techniques, specifically deep learning, allows for solving multi-level problems associated with recognizing text in an image. Using this approach, it is possible to work with a much larger set of input data, since their characteristics are taken into account at each stage of the functioning of neural networks and have little impact on the result, provided that the model is highly trained.

Figure 7 – Convolutional recurrent neural network

Results and discussion

As an example of one of the modern OCR systems, Tesseract should be mentioned. This mechanism was originally developed by the Hewlett-Packard Laboratories research group in the 1980s and released as open-source functionality in 2005 and at that time the system was only capable of recognizing text in English [21]. Currently, this system provides the ability to work with several languages simultaneously. The general principle of Tesseract is to analyze an image and discover patterns for character recognition. The first step involves pre-processing the image, including increasing contrast and reducing noise, then executing the algorithms provided by the engine for feature detection and extraction, character edge detection, and pattern matching [22]. In its operation, Tesseract also utilizes deep learning techniques, such as convolutional neural networks, as well as a variant of recurrent neural networks called long short-term memory (LSTM) [23]. Through this approach, it is possible to recognize text in different languages and under different conditions, for example, when recognizing handwritten text [24].

Tesseract can be integrated into various functionality related to image recognition through the API and used in code written in programming languages such as C#, Python, C++ and so on [25]. For example, in the context of a Windows Forms application, in order to work with this system, it is necessary to connect the functionality to the working environment and initialize the mechanism by creating a new instance of the TesseractEngine class indicating the necessary parameters (language, mode, and so on). After this, the created object is used to access the Tesseract programming interface, in particular, to the Process() method, which is responsible for processing a pre-prepared image presented in a machine-readable format (Figure 8).

```
// Tesseract initialization
using (var engine = new TesseractEngine(@"C:                                    \Tesseract-OCR
  \tessdata", "eng+rus", EngineMode.Default))
{
    // Increasing the contrast of drawn lines
    Bitmap highContrastBitmap = PreprocessImage(drawingBitmap);

    // Saving a temporary image file
    temporaryImageFile = Path.GetTempFileName();
    highContrastBitmap.Save(temporaryImageFile, System.Drawing.Imaging.ImageFormat.Jpeg);

    // Loading the temporary image file using Tesseract
    using (var image = Pix.LoadFromFile(temporaryImageFile))
    {
        using (var page = engine.Process(image))
        {
            string recognizedText = page.GetText();
            MessageBox.Show($"Recognized text: {recognizedText}");
            textBox1.Text = recognizedText;
        }
    }
```

Figure 8 – Initialization of Tesseract and execution of recognition

Thus, Tesseract, using its built-in optical character recognition techniques, is capable of processing images and extracting even handwritten characters in various languages. Figure 9 illustrates the result of processing an image with handwritten text in English.

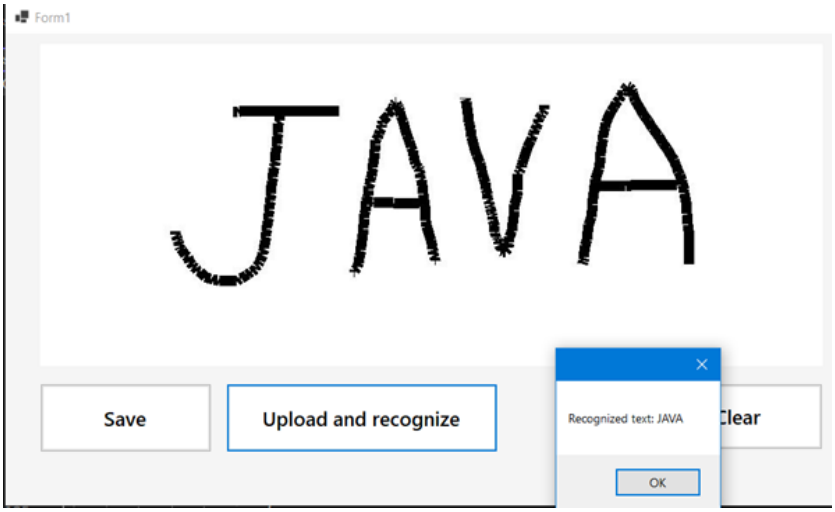

Figure 9 – The result of the algorithm execution with the text in English

The result of the mechanism processing an image with handwritten text in Russian is demonstrated in Figure 10.

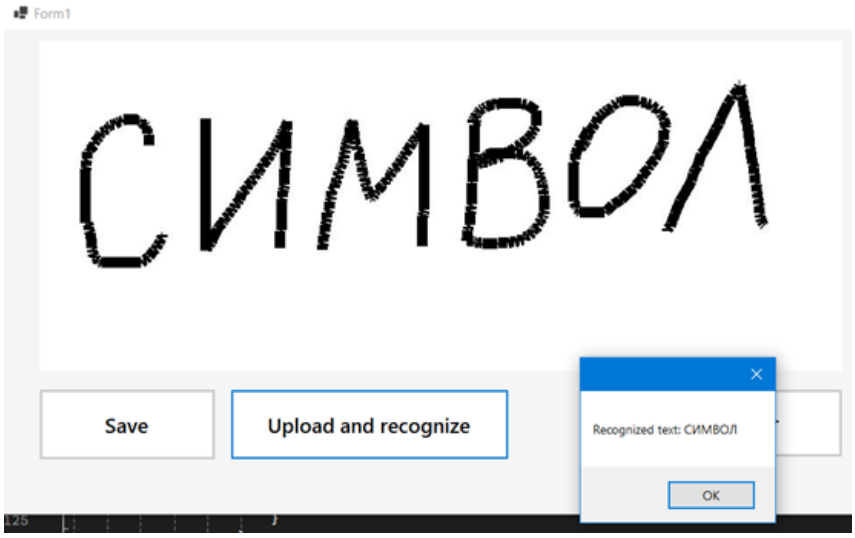

Figure 10 – The result of the algorithm execution with text in Russian

Conclusion

Thus, machine learning techniques, in particular deep learning techniques, are regularly applied in the context of optical character recognition technology, as they enable the automatic extraction of characteristic features from raw data and generate a high-quality result based on these features in the form of correctly converted text in the image. These approaches and OCR technology itself continue to develop and find their application in various situations, ensuring efficiency, accessibility, and automation of business processes. The choice of neural network architecture depends on the specifics of the task, the available data, the required level of accuracy, as well as the data processing speed.

REFERENCES

1 Singh A., Bacchuwar K., Bhasin A. A survey of OCR Applications. International Journal of Machine Learning and Computing, 2012, vol. 2, no. 3, pp. 314–318.

2 Shruthi P. Verma D. C. A Detailed study and recent research on OCR. International Journal of Computer Science and Information Security, 2021, vol. 19, no. 2, pp. 52–66.

3 Ahmed. M, Abidi A.I. Review on optical character recognition. International Research Journal of Engineering and Technology (IRJET), 2019, June, v0l. 06, issue 06, pp. 3666–3669.

4 Verma P., Foomani G. M. Improvement in OCR Technologies in Postal Industry Using CNN-RNN Architecture: Literature Review. International Journal of Machine Learning and Computing, 2022, vol. 12, no. 5, 154–163.

5 What is optical character recognition (OCR)? Ibm.com. January 5, 2022. https://www.ibm.com/blog/optical-character-recognition/

6 Modi H., Parikh M. C. A Review on Optical Character Recognition Techniques International Journal of Computer Applications, 2017, vol. 160, no. 6, pp. 20–24.

7 Fateh A., Fateh M., Abolghasemi V. Enhancing optical character recognition: Efficient techniques for document layout analysis and text line detection. Engineering Reports, 2023, November, pp. 1–26.

8 Wang X.F., He Z.H., Wang K., Wang Y.F., Zou L. A survey of text detection and recognition algorithms based on deep learning technology. Neurocomputing, 2023, November, vol. 556.

9 AI vs. Machine Learning vs. DeepLearning vs. Neural Networks: What's the difference? Ibm.com. July 6, 2023. https://www.ibm.com/blog/ai-vs-machine-learning-vs-deep-learning-vs-neural-networks/

10 Subedi B., Yunusov J., Gaybulayev A., Kim T. Development of a Low-cost Industrial OCR System with an End-to-end Deep Learning Technology. IEMEK J. Embed. Sys. Appl, 2020, April, pp. 51–60.

11 Difference Between Machine Learning and Deep Learning. GeeksForGeeks.org. June 5, 2023. https://www.geeksforgeeks.org/difference-between-machine-learning-and-deep-learning/

12 Meng F., Ghena B. Research on Text Recognition Methods Based on Artificial Intelligence and Machine Learning. Advances in Computer and Communications, 2023, November, pp. 340–344.

13 Memon J., Sami M., Khan R. A. Handwritten Optical Character Recognition (OCR): A Comprehensive Systematic Literature Review (SLR). IEEE Access, 2020, July, pp. 142642–142668.

14 Rakesh S., Reddy P.K., Prashanth V., Reddy K.S. Reddy. Handwritten text recognition using deep learning techniques: a survey. ICMED, 2024, pp. 1–8.

15 Nikolenko S., Kadurin A., Arkhangelskaya E. Glyubokoye obucheniye: Seriya «Biblioteka programmista» (SPb.: Izd-vo Piter, 2018), pp. 231–259 [in Russian].

16 Hemanth G.R., Jayasree M., Keerthi Venii S., Akshaya P., and R. Saranya. CNN-RNN based handwritten text recognition. ICTACT Journal on Soft Computing, 2021, October, vol. 12, pp. 2457–2463.

17 Ahlawat S., Choudhary A., Nayyar A., Singh S., Yoon B. Improved Handwritten Digit Recognition Using Convolutional Neural Networks (CNN). Sensors, 2020, June, pp. 1–18.

18 Goodfellow I, Bengio Y, Courville A. Deep Learning, (The MIT Press, 2015), pp. 297–329.

19 Shinde S., Saraiya T. Using CRNN to Perform OCR over Forms. International Journal of Engineering Research & Technology (IJERT), vol. 9, pp. 319–323.

20 Shi B., Bai X., & Yao C. An end-to-end trainable neural network for image-based sequence recognition and its application to scene text recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016, pp. 2–3.

21 Clausner C., Antonacopoulos A., Pletschacher S. Efficient and effective OCR engine training. International Journal on Document Analysis and Recognition (IJDAR), 2019, October, vol. 23, pp. 73–88.

22 Mursari L.R., Wibowo A. The Effectiveness of Image Preprocessing on Digital Handwritten Scripts Recognition with The Implementation of OCR Tesseract. Computer Engineering and Applications, 2021, October, vol. 10, pp. 177–186.

23 Bagwe S., Shah V., Chauhan J., Harniya P., Tiwari A., Gupta V., Raikar D., Gada V., Bheda U., Mehta V., Warang M., Mehendale N. Optical character recognition using deep learning techniques for printed and handwritten documents. SSRN Electronic Journal, 2020, January, pp. 1–10.

24 Patel C., Patel A., Patel D. Optical Character Recognition by Open Source OCR Tool Tesseract: A Case Study. International Journal of Computer Applications, 2012, October, vol. 55, pp. 50–56.

25 Jain P., Dr. Taneja K., Dr. Taneja Harmunish. Which OCR toolset is good and why? A comparative study. Kuwait J.Sci., 2021, April, vol. 48 (2), pp. 1–12.

[1]Выходцева В.А.,
магистрант, ORCID ID: 0009-0007-1897-0768
e-mail: vykhodtseva.va@gmail.com
[1]Попова Г.В.,
ф–м.ғ.к., қауымдастырылған профессор, ORCID ID: 0000-0002-6935-1066
e-mail: gpopova@edu.ektu.kz

[1]Қазақстан-Американдық еркін университеті,
070000, Өскемен қ., Қазақстан

ТАҢБАЛАРДЫ ОПТИКАЛЫҚ ТАНУ НӘТИЖЕЛЕРІНІҢ ДӘЛДІК ДЕҢГЕЙІН АРТТЫРУ ҮШІН МАШИНАЛЫҚ ОҚЫТУ ӘДІСТЕРІН ҚОЛДАНУ

Аңдатпа

Қазіргі заманның ең кең таралған үрдістерінің бірі – адамзат өмірінің барлық негізгі салаларын қамтыған цифрландыру. Ақпараттық технологиялардың дамуы күнделікті өмірдің ғана емес, сондай-ақ кәсіпкерлік, экономика, денсаулық сақтау саласындағы күрделі бизнес-үрдістерді автоматтандыру арқылы жаһандық деңгейдегі өзгерістерге де ықпал етті. Цифрлық деректер мен құжаттамаға көшу қажетті ақпараттың қолжетімділігін қамтамасыз етіп қана қоймай, оны талдау мен өңдеудің тиімділігін арттырды. Осыған байланысты мәтіндік деректерді суреттерден анықтауға және алуға мүмкіндік беретін оптикалық таңбаларды тану (OCR) технологиялары маңызды. OCR технологиялары қоғамның цифрлық трансформациясында шешуші рөл атқарады, өйткені олар суреттердегі мәтіндік ақпаратпен қолмен жұмыс істеу қажеттілігін жояды және қағаз тасымалдағыштардағы деректерді өңдеуге қатысты көптеген бизнес-процестерді автоматтандыруға мүмкіндік береді. Мысалы, қағаз нысандарынан статистикалық деректерді жинау, қағаз құжаттарын электрондық құжат айналымы жүйесіне енгізу, мәтіндік ақпаратты аудио файлдарға түрлендіру сияқты процестерде қолданылады. Бұл мақала оптикалық таңбаларды тану технологиясын сипаттауға және алынған нәтижелердің сапасын жақсарту мақсатында оны заманауи іске асыру аясында белсенді қолданылатын машиналық оқыту әдістеріне шолу жасауға арналған. Сонымен қатар мақалада сипатталған тәсілдердің жұмыс принциптері, олардың мүмкіндіктері, сондай-ақ белгілі бір сценарийлерде қолдану кезінде кездесетін кейбір шектеулер қарастырылған.

Тірек сөздер: таңбаларды оптикалық тану, белгілер, машиналық оқыту, терең оқыту, нейрондық торап желілері, қайталанатын нейрондық желілер.

[1]Выходцева В.А.,
магистрант, ORCID ID: 0009-0007-1897-0768
e-mail: vykhodtseva.va@gmail.com
[1]Попова Г.В.,
к.ф–м.н., ассоциированный профессор, ORCID ID: 0000-0002-6935-1066
e-mail: gpopova@edu.ektu.kz

[1]Казахстанско-Американский свободный университет,
070000, г. Усть-Каменогорск, Казахстан

ПРИМЕНЕНИЕ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ПОВЫШЕНИЯ УРОВНЯ ТОЧНОСТИ РЕЗУЛЬТАТОВ ОПТИЧЕСКОГО РАСПОЗНАВАНИЯ СИМВОЛОВ

Аннотация

Одним из самых широко распространенных процессов современности, безусловно, является цифровизация, которая охватила все ключевые сферы жизни человечества. Развитие информационных технологий поспособствовало масштабным изменениям не только в повседневном аспекте жизнедеятельности, но и в

более глобальном, автоматизировав сложные бизнес-процессы в сфере предпринимательства, экономики, здравоохранения. Переход к цифровым данным и документации обеспечил большую доступность необходимой информации, а также повысил эффективность ее анализа и обработки. В связи с данным фактом важное значение обрели технологии оптического распознавания символов (OCR), позволяющие определять и извлекать текстовые данные из изображений. OCR-технологии играют ключевую роль в цифровой трансформации общества, поскольку они исключают необходимость ручной работы с текстовой информацией на изображениях и применимы в автоматизации большинства бизнес-процессов, связанных с обработкой данных на бумажных носителях, например, при сборе статистических данных из бумажных форм, отражении бумажных документов в системе электронного документооборота, конвертации текстовой информации в аудиофайлы и так далее. Данная статья посвящена описанию технологии оптического распознавания символов, а также обзору методов машинного обучения, которые активно применяются в контексте ее современной реализации с целью улучшения качества получаемых результатов. Кроме того, в статье представлены принципы работы описываемых подходов, их возможности, а также некоторые ограничения, с которыми можно столкнуться при их использовании в тех или иных сценариях.

Ключевые слова: оптическое распознавание символов, признаки, машинное обучение, глубокое обучение, сверточные нейронные сети, рекуррентные нейронные сети.

Article submission date: 28.05.2024.