IRSTI 28.23.02

Aisaule Bazarkulova[1], Yelnur Mutaliyev[2], Abylaikhan Chazhabayev[3], Duman Telman[4], Diana Bazarkulova[5]
[1,2,3,4,5]Suleyman Demirel University, Kaskelen, Kazakhstan

KAZAKH HANDWRITING RECOGNITION

Abstract. Recognition of handwritten text is one aspect of object recognition and known as handwriting detection cause of a computer's potential to recognize and comprehend readable handwriting from resources including paper files, touch smart devices, images, etc. Data is categorized into a number of classes or groups using pattern recognition. The paper presents a successful experiment in recognizing handwritten Kazakh text using Convolutional Recurrent Neural Network based architectures and the Kazakh Autonomous Handwritten Text Dataset. The proposed algorithm achieved an overall accuracy of 86.36% and showed promising results. However, the paper suggests that further research could be conducted to improve the model, such as correlating and enlarging the database or incorporating other models and libraries. Additionally, the paper emphasizes the importance of considering language specifics when building a text recognition model, as modern algorithms that work well in one language may not guarantee the same performance in another.

Keywords: recognition, handwrite detection, Kazakh language, binarization method

***

Аңдатпа.Қолжазба мәтінін тану нысанды танудың бір аспектісі болып табылады және қолжазбаны тану деп аталады, өйткені компьютер ресурстардан оқылатын қолжазба мәтінін, соның ішінде қағаз файлдарын, сенсорлық ақылды құрылғыларды, суреттерді және т.б. тануға және түсінуге қабілетті. Мақалада конволюциялық қайталанатын нейрондық желілер және қазақ тіліндегі дербес қолжазба мәтінінің деректер жиынтығы негізінде архитектураларды пайдалана отырып, қазақ тіліндегі қолжазба мәтінді тану бойынша сәтті эксперимент ұсынылған. Ұсынылған алгоритм жалпы дәлдікке 86,36% жетті және перспективалы нәтижелер көрсетті. Дегенмен, құжат дерекқорды салыстыру және кеңейту немесе басқа модельдер мен кітапханаларды қосу сияқты үлгіні жақсарту үшін қосымша зерттеулер жүргізуді ұсынады. Сонымен қатар, мақалада мәтінді тану моделін құру кезінде тілдік ерекшеліктерді ескерудің маңыздылығы атап өтіледі, өйткені бір тілде жақсы жұмыс істейтін заманауи Алгоритмдер екінші тілде бірдей өнімділікке кепілдік бермеуі мүмкін.

Түйінді сөздер: қолжазба мәтінді тану, тану, қазақ тілі, бинаризация әдісі.

***

Абстракт. Распознавание рукописного текста является одним из аспектов распознавания объектов и известно как распознавание рукописного ввода, поскольку компьютер способен распознавать и понимать читаемый рукописный текст из ресурсов, включая бумажные файлы, сенсорные интеллектуальные устройства, изображения и т.д. Данные подразделяются на несколько классов или групп с помощью распознавания образов. В статье представлен успешный эксперимент по распознаванию рукописного текста на казахском языке с использованием архитектур на основе сверточных рекуррентных нейронных сетей и набора данных автономного рукописного текста на казахском языке. Предложенный алгоритм достиг общей точности 86,36% и показал многообещающие результаты. Однако в документе предлагается провести дальнейшие исследования для улучшения модели, такие как сопоставление и расширение базы данных или включение других моделей и библиотек. Кроме того, в статье подчеркивается важность учета языковых особенностей при построении модели распознавания текста, поскольку современные алгоритмы, которые хорошо работают на одном языке, могут не гарантировать такую же производительность на другом.

Ключевые слова: распознавание, распознавание рукописного текста, казахский язык, метод бинаризации.

Introduction

Today, many systems have been developed that demonstrate the capabilities of artificial neural networks, for example, networks capable of representing text phonetically, recognizing handwritten letters. Now keyboard input is replacing handwritten text, but for many years handwritten text has been used in documentation and communication in society. But, nevertheless, this tool remains the simplest and most effective way in terms of time to transform human thought into a transferable form for most people around the world. After the development of information technology, the transformation of handwritten text into computer-understandable data has become an urgent task today. Research has mainly focused on the recognition of the Latin languages' handwriting. Only fewer studies have been done for the Kazakh language. It is rare to find algorithms for recognizing the Kazakh language, because this will greatly simplify the task in many areas of activity, where there is documentation in the Kazakh language.

The paper of [1] uses convolutional neural networks to recognize Arabic characters (CNN). In terms of big data and visuals, CNN performs better than other deep learning models. They utilize CNN's strength in managing huge dimensions of input and sharing weights. In an experimental section they showed that the results were promising with a 94.9% classification accuracy rate on testing images.

At this stage of development, handwriting recognition systems have more and more new horizons for advances in such areas as machine translation, mail forwarding, signature verification, daily note taking, etc. The ultimate goal of all handwriting recognition systems is to create machines that can read any text with the same recognition accuracy as people, but at the same time with greater speed. The handwriting recognition system handles the formatting, performs proper character segmentation, and finds the most plausible words.

Postal addresses, bank check amounts and forms are the vivid examples of handwriting recognition. That is why handwriting recognition plays an important role in digital world.

First project objective is to recognize handwritten documents, which includes characters, words, lines, numbers, paragraphs, etc. The second is to apply handwriting recognition in many fields, such as banks, hospitals, government agencies, etc. Given the popularity of smartphones and tablets, where writing with a finger or a stylus is anticipated to be a potentially effective method of input, handwritten text recognition is still a significant issue that is reviving interest as a current field of study. The format in which text is given for recognition greatly affects how complex the work of text recognition is. It should be mentioned that the printed text is distinguished by the fact that it is consistently positioned on the page in even lines, and the text's characters typically have the same height and breadth throughout the concerned document. Furthermore, the space between the characters is easily discernible and frequently has the same width throughout a sentence. As a result, the intricacy of text recognition is made simpler by the inclusion of these characteristics. The challenge of handwritten text recognition is currently not entirely accomplished. This makes the topic's consideration pertinent in this regard.

We propose a novel approach for manuscript recognition that addresses the challenges of recognizing handwritten text. Our approach involves breaking the text into segments and using a neural network consisting of three layers for recognition. We achieved state-of-the-art results on a benchmark dataset of handwritten text in Kazakh, which has implications for document analysis, preservation, and digitization. Our contribution is significant because it can potentially improve the accuracy and efficiency of manuscript recognition for various languages.

A digitized view of handwritten records would allow automating the business processes of many companies, simplifying the work of a person. It will be used by banks to process checks, and by the postal service to recognize addresses.

The rest of the paper is organized as follows. Next section gives an overview of Kazakh alphabet and its handwriting versions. Following section discusses the difference between the existing handwriting detection methods. Section 3 discusses the methodology of the research and used data set. Section 4 presents the way how dataset was prepared and introduces to the model itself. It shows the solution with the neural networks and try-out on the datasets. Section 6 shares the results of the detection and analysis of each try. Section 7 concludes with the summary and experience of applying this technology.

Literature Review

Since Kazakh language currently uses Cyrillic alphabet we were quite interested in the solution of the other handwriting recognition problem which also use Cyrillic characters. It is possible to create and use CNN algorithms to recognise characters of Mongolian alphabet [2]. The author says that LeNet-5 architecture has an appropriate structure for an image recognition. There are two applied functions of Convolutional and SubSampling layers, which describe the network model. Their experiment showed good results having approximately 3% of error rate after 1000 passes.

The handwriting recognition surrounds a big range of important processes problems, starting from speech recognition and the classification of handwritten characters. Author presents the state-of-art in handwriting recognition by exploring trends , made use of four approaches for handwriting recognition : template matching, statistical classification, structural matching and neural Networks [3]. The advantage of the paper is researcher utilises the image preprocessing / cleaning up the images as the first step. Which will be used for the feature extracted component's efficient result. This step is used in order to convert the paper based documents. Author uses Neural Networks to analysis. Neural Networks prepare the data to the future analysis of recognition.

In this paper authors collected and analyzed the topic of handwritten , which were published between year 2000 to 2018 [4]. Researches in purpose to solve the problem , which is recognition of handwriting , used different types of methods as a : artificial neural networks, kernel method and statistical method. As the advantages , authors included in the page recognition search strategy. Because of the fact that it has automatic and manual search the result of efficient and convenient work. Second advantage is they included enough of dataset of many different languages for testing and not only. As the disadvantages , researches performed parametric classifiers. Parametric classifiers have limit of input data size which is not convenient, because recognition uses big data. Conclusion of the paper is authors successfully extracted and analyzed research publications on six widely spoken languages.

[5] used the HTR formula, that consists of 3 main components: the rummage around for the Start-offline finder, Line Follower, and handwriting recognizer. Their 1st mission is to seek out the beginning points of every text line. The Start-of-Line (SOL) follower is Associate in Nursing RPN with a truncated terribly Deep Convolutional Network VGG-19 design. Once the start line of every line is set mistreatment SOL, subsequent task is to follow the line wherever the text seems. For this purpose, a fraction of the image is cut out at the start of the line. Given the SOL and full line on that the text is found, the ultimate task is to acknowledge written characters. For this purpose, the CNN-LSTM HWR deep learning design is taken into account. The advantage of the article is improving the linear sequence a part of their projected formula will improve performance. The disadvantage of the article is their projected formula tries to predict individual characters, not words.

[6] used a replacement neural network model of AI to acknowledge handwriting and to acknowledge a man or woman because the author of a written text. All signs of little letters were combined into one abstract letter (some quite symbol). They labelled it as S. They used one logical letter S for all little letter functions. The recognition system consists of 2 subsystems: the primary neural network and also the second neural network. Advantages of this article: The problem of recognizing written characters was investigated; A neural network model of AI. Disadvantages of this article: This misclassification error can not be but 10-4; Special functions can not be extracted mechanically.

At the moment, one of the most frequently used platforms for handwriting recognition is systems from Google (“Google Translate”, “Google Handwriting Input”). The article from [7] ”Multi-Language Online Handwriting Recognition” describes Google systems that currently support recognition of 97 languages. The authors show that the use of elements depends on the language that is recognized, but also some components can be used for other languages a secondary time, which is the advantages of this system. As new ideas, the authors present the recognition of overlapping texts and the processing of diacritics using joint decoding, which is interpreted from the input data based on time and position into a single lattice. The system also accepts various inputs from touch-screen mobile phones and mouse-drawn handwriting, which is one of the advantages.

Since one of our goals is to cover more areas where there is a handwriting. One of the materials where the handwriting is written is a whiteboard, which is used in many fields of activity. In the article of [8], they consider the recognition of manuscripts made on a whiteboard. The article used Recurrent Neural Network(RNN) for recognition, or rather its new Connectionist Temporal Classification (CTC) function, which in sequence gives the marking of non-segmented data. Function as a result gives a handwritten text recognition rate of 74%, which is 8.6% more than the results based on the Hidden Markov Model (HMM).

As handwriting recognition was one of the difficult and interesting research areas, this paper by [9], explained and described some methods of offline and online handwritten alphabetical character recognition systems. The representative handwriting recognition system works within pre-processing and segmentation stages, proposed feature extraction, classification and recognition, and post-processing stages . And in the article mentioned proposed feature extraction method, presented for extracting the features of alphabet, which is important as its efficient operation raises recognition rates and lowers misclassification. The suggested handwrite recognizer will have high recognition rates and will be ideally suited for many applications, such as document reading and the translation of any handwritten material into structural text.

Some of the interesting articles about classification and deep learning models in order to solve the problem of the HTR says that classification could be based on CNN and three models such as SimpleHTR, Bluche, Puigcerver [10]. It showed good results of experiments that provided on various datasets. Each model contains different layers and parts.

Within the article from [11], said that handwritten recognition is one type of optical character recognition (OCR). OCR goes within the text, which can be printed text either handwritten. They explain, as the name implies, printed character recognition is the identification of characters in an image of a newspaper or other printed document. Characters written by humans or influenced by humans are recognized when they are written in handwriting. There are many ways to identify handwriting as said in the article. Convolutional Neural Network (CNN), Zoning, incremental and semi-incremental segmentation, as well as slope and slant correction, are a few of these. The Slope and Slant Correction method has the lowest accuracy of these techniques, whereas Convolutional neural networks (CNN) have the highest accuracy. This is one of effective techniques of identifying handwriting, and when the CNN algorithm is trained on the images, we will get decent validity. This method's biggest flaw is that it takes too long to train the model because so many image samples are needed.

For our solution, as CNN, one of the more modern methods for recognition compared to OCR, is our chosen approach for recognition. OCR is frequently utilized when reading papers regarding Kazakh text recognition. Because we obtain good accuracy when using CNN to train images, this is one of the successful ways for handwriting recognition. CNNs, which are completely interconnected neural networks with direct communication, are exceptionally good at reducing the number of parameters without sacrificing the quality of the model. Since each pixel is taken into account as a separate function, the images have a high dimension, which is consistent with the previously mentioned CNN capabilities.

Methodology

Handwritten character recognition is considered to be in the field of artificial intelligence. It is an ongoing field of research that mostly requires data driven methodology. This methodology proposes an effective approach to use collected data or ready dataset to obtain relevant information for analysis and interpretation. For our research, it is an ideally suited principal reason to choose this methodology that creates conditions for collating with previous results and suggesting an ameliorated and advanced way to figure out which model and its techniques are the best. We used an extensive and well-organized dataset to accomplish our main objectives.

The dataset that we are going to use for our research is KOHTD [12], which was collected relatively recently (there was no dataset before). Although KOHTD does not contain all the words that we have (at the moment it is unrealistic to collect all the words in written form), it contains works that are written in Cyrillic and showing all 42 letters that are in our alphabet. Dataset was collected from works of students of universities of our country (Satbayev Universality and Al-Farabi Kazakh National University), also 99% of works are written in Kazakh language and remaining 1% is Russian language from 3000 answers on exam.

Next, we break the numbers into segments. Thus, the text is initially divided into numbers representing individual segments, then each segment is analyzed, and individual segment symbols are highlighted in it. After that we will use an NS of three layers:

- The input level of the network contains neurons garbling colorful values of input pixels.
- The alternate level of the network is hidden. We'll denote the number of neurons in this level n.
- There are 10 neurons in the output level of the network. If the first neuron is activated, that is, its output value is = 1, this indicates that the network believes that the input was 0. If the second neuron is activated, the network considers that there was 1 at the input. Strictly speaking, we number the output neurons from 0 to 9, and look at which of them had the maximum activation value.

Regarding the evaluation of our Convolutional Neural Network (CNN) model's performance, we evaluated it using the standard accuracy metric, which is the percentage of correctly classified instances out of the total number of instances in the test set. In other words, it measures how often the model's prediction matches the true label of the input. The accuracy can be calculated using the following formula:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

where:

- **TP** is the number of true positives (i.e., the number of instances that are correctly classified as positive).
- **TN** is the number of true negatives (i.e., the number of instances that are correctly classified as negative).
- **FP** is the number of false positives (i.e., the number of instances that are incorrectly classified as positive).
- **FN** is the number of false negatives (i.e., the number of instances that are incorrectly classified as negative).

In addition to accuracy, other commonly used metrics for evaluating classification models include precision, recall, and F1 score. Precision measures the proportion of positive predictions that are true positives, while recall measures the proportion of true positives that are correctly predicted as positive. The F1 score is a weighted average of precision and recall, and is often used as a single metric to evaluate a model's overall performance.

It's worth noting that the choice of evaluation metric can depend on the specific task and goals of the model. For example, in some cases, minimizing false positives (i.e., maximizing precision) may be more important than maximizing overall accuracy.

Sure, the character error rate (CER) is another commonly used metric for evaluating CNN systems. It measures the percentage of characters in

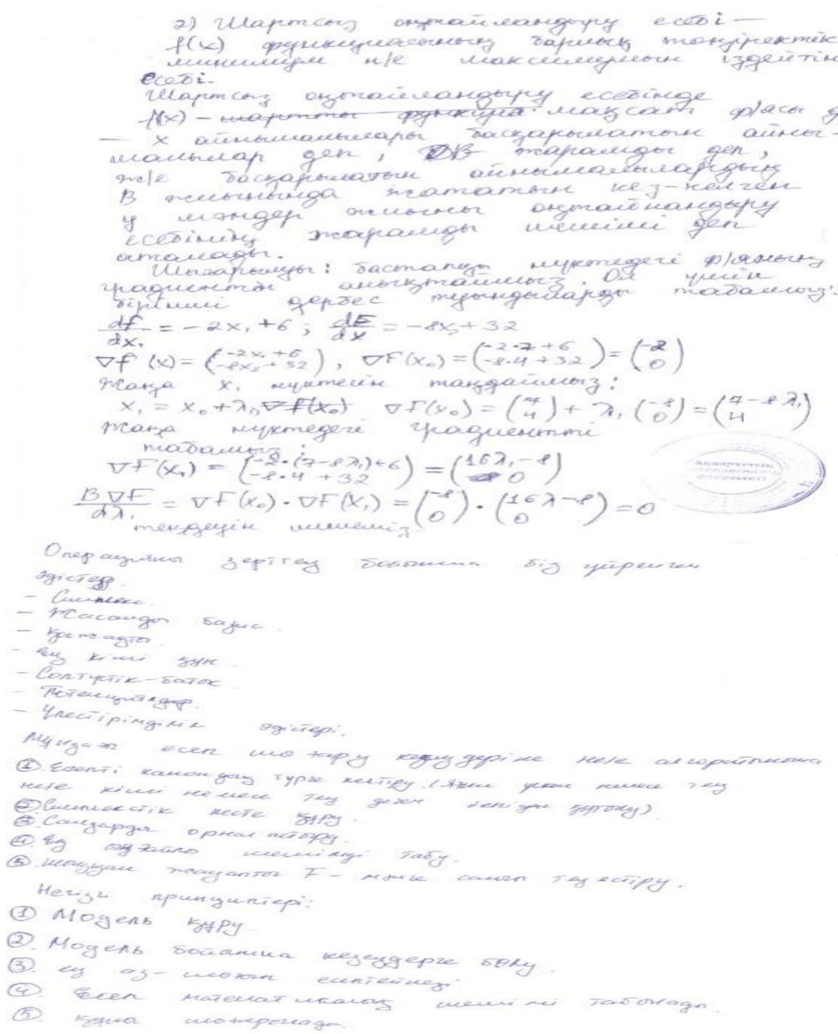


Figure 1: Samples from the dataset

the transcription that is incorrect, regardless of their position in the word. CER can be calculated using the following formula:

$$CER = \frac{SC + DC + IC}{GT} \tag{2}$$

where:

- **SC** is the number of substitutions (i.e., the number of characters in the OCR output that differ from the ground truth).
- **DC** is the number of deletions (i.e., the number of characters in the ground truth that are missing from the OCR output).
- **IC** is the number of insertions (i.e., the number of extra characters in the OCR output that are not in the ground truth).
- **GT** is the total number of characters in the ground truth.

Word error rate (WER) is similar to CER, but it measures the percentage of words in the transcription that are incorrect, regardless of their position in the sentence. WER is calculated in the same way as CER, but at the word level instead of the character level.

Main Part

Data Preparation

Reducing Noises

In most cases, the raw data that were collected contains noise, that is, unwanted features that make it difficult to perceive the image. Although these images can be used directly to extract features, the accuracy of the algorithm will suffer greatly. To remove noise, we used OpenCV Morphological Operations - this is one of the image processing methods that processes an image based on a shape. The opening involves erosion followed by expansion of the outer surface (foreground) of the image. Closure involves expansion, followed by erosion of the outer surface (foreground) of the image. They are used to remove noise in the image.

Binarization

A binarization method is the interpretation of a color (or grayscale) picture into two–color black and white. The fundamental parameter of the sort of transformation is the brink t – price with which the brightness of every pixel is compared. For example, if the Threshold(t) value is 240, then all pixels with values greater than 240 will be assigned the value 1, that is the pixel will be considered white, and all pixels with values less than or equal to this will be assigned the value 0, that is it becomes black.

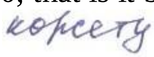
a)  b) 

Figure 2: a) original image b) image after binarization

Thinning and Skeletonization

Thinning and Skeletonization allow us to reduce or increase the stroke width of the object if it is necessary. Since we have used the KOHTD dataset [12], where each writer’s handwritten text has a different stroke width and handwriting style. To perform this, we have used only two basic morphological operations: dilate and erode.

- Dilation is an operation in which white areas inside an image “grow“ or ”dilate”.
- Erosion is an operation in which white areas inside an image are “shrink” or “erode”.

As a result, all images were obtained with approximately the same stroke volume, which will be more convenient to work with.

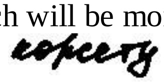
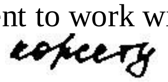
a)  b) 

Figure 3: a) binarized image b) image after thinning

Preparing Validation Data

To train data we need some information about each image to validate. In our case, they are the names of image files and their labels. Since this dataset stores each information in a separate JSON file, we had to collect them into one CSV file [13]. These are our dependent variables which are going to be predicted by proposed model, this includes descriptions against our independent variables, we need to define our dependent variable while training the model.

	name	description
0	0_002_001.jpg	1
1	0_002_002.jpg	1С:
2	0_002_003.jpg	Кәсіпорын
3	0_002_004.jpg	жүйесіндегі
4	0_002_005.jpg	еңбекақыны,

Figure 4: Pandas DataFrame of collected JSON files

Proposed CRNN Model and Its Architecture

The model used for the problem is called Convolutional Recurrent Neural Network because it integrates Deep Convolutional Neural Networks and RNN. This model consists of three components where there are convolutional layers, recurrent layers, and transcription layer. Our model is similar to [14] model, but it has some slightly changes. It was used CNN, Bi-directional LSTM for RNN trained using Adam Optimizer [15].

Look at the Figure 5 to see details of the Model Architecture:

```
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
input (InputLayer)           [(None, 32, 128, 1)]      0

conv1 (Conv2D)               (None, 32, 128, 64)       640

batch_normalization_72 (Bat  (None, 32, 128, 64)       256
chNormalization)

activation_68 (Activation)   (None, 32, 128, 64)       0

max1 (MaxPooling2D)          (None, 16, 64, 64)        0

conv2 (Conv2D)               (None, 16, 64, 128)       73856

batch_normalization_73 (Bat  (None, 16, 64, 128)       512
chNormalization)

activation_69 (Activation)   (None, 16, 64, 128)       0

max2 (MaxPooling2D)          (None, 8, 32, 128)        0

dropout_46 (Dropout)         (None, 8, 32, 128)        0

conv3 (Conv2D)               (None, 8, 32, 256)        295168

conv4 (Conv2D)               (None, 8, 32, 256)        590080

batch_normalization_74 (Bat  (None, 8, 32, 256)        1024
chNormalization)

activation_70 (Activation)   (None, 8, 32, 256)        0

max3 (MaxPooling2D)          (None, 8, 16, 256)        0

conv5 (Conv2D)               (None, 8, 16, 512)        1180160

batch_normalization_75 (Bat  (None, 8, 16, 512)        2048
chNormalization)

dropout_47 (Dropout)         (None, 8, 16, 512)        0

reshape (Reshape)            (None, 64, 1024)          0

dense1 (Dense)               (None, 64, 64)            65600

lstm1 (Bidirectional)        (None, 64, 512)           657408

lstm2 (Bidirectional)        (None, 64, 512)           1574912

dense2 (Dense)               (None, 64, 87)            44631

softmax (Activation)         (None, 64, 87)            0

=================================================================
Total params: 4,486,295
Trainable params: 4,484,375
Non-trainable params: 1,920
```

Figure 5: Model Architecture

1 Input shape: The input shape of the image is specified as (32, 128), where 32 is the height and 128 is the width of the image.
2 Convolution layers: The model includes five convolution layers, with each layer having a kernel size of (3, 3). The number of filters used in these layers starts at 64 and increases up to 512 in the later layers. The selection of the number of filters in each convolution layer is a hyperparameter, and can impact the model’s performance.
3 MaxPooling layers: Three MaxPooling layers are added to the model, with two of them having a size of (2, 2) and the third having a size of (1, 2). The size of the MaxPooling layer is a hyperparameter that determines the size of the pool window used for downsampling.

4 Batch normalization: Batch normalization layers are included in the model, which help to accelerate the training process by normalizing the inputs to each layer. The hyperparameter used in batch normalization is the momentum, which controls the moving average of the mean and standard deviation used for normalization.

5 Bidirectional LSTM layers: The model includes two Bidirectional LSTM layers, each with 128 units. The number of units in the LSTM layer is a hyperparameter that determines the capacity of the network to learn complex patterns and relationships in the input data.

The model can be evaluated using various performance metrics, such as:

Result and Analysis

In this section we would like to sum up the results. For evaluation purposes, we have prepared labels for CTC Loss. The labels were converted to numbers which represent each character in the train data that contain Kazakh alphabet and symbol characters.

```
True label :  болады .
train_y :  [43. 59. 55. 41. 47. 76. 82. 83. -1. -1. -1. -1. -1. -1. -1. -1. -1. -1.
 -1. -1. -1. -1. -1. -1. -1. -1. -1. -1. -1. -1.]
train_label_len :  [8.]
train_input_len :  [30.]
```

Figure 6: Example of encoding

The machine’s processor is AMD Ryzen 7 5800H with Radeon Graphics and graphics card is NVIDIA GeForce RTX 3050 Ti with 16 GB of RAM. The training parameters were set to values up to 50 epoch, batch size of 128 in the image and learning rate of 0.001.

The model accuracy shows good result with the 86.36% accuracy, which is better than [13] model performance. 93.73% of characters predicted correctly which is also satisfying result.

The accuracy of the handwritten text recognition model was computed by comparing the predicted output of the model with the actual output. In our case, the actual output would be the ground truth text for the input image. The predicted output would be the text recognized by the model for the given input image.

```
Characters predicted correctly :  93.79%
Words predicted correctly      :  86.36%
```

Figure 7: Results

Looking at the results of training and validation we make a summary that when number of epochs growing the accuracy shown better results (Figure 8).

Conclusion and Discussion

In this paper, we examined the recognition of handwritten Kazakh text. To achieve this goal, we developed and used the state-of-the-art Convolutional Recurrent Neural Network based architectures. To teach the model and check its accuracy we used Kazakh Autonomous Handwritten Text Dataset [12], which consists of a large collection of exam papers filled out by students of Satbayev University and AlFarabi Kazakh National University. The experiment shows that the proposed algorithm has 86.36% of predicted word correctly overall. The model showed good results. In the future, it is possible to conduct research to improve the compiled trained model. This can be done by further correlating the database and enlarging the existing database. We can also use other well-known models to train and add additional libraries to improve the algorithm.

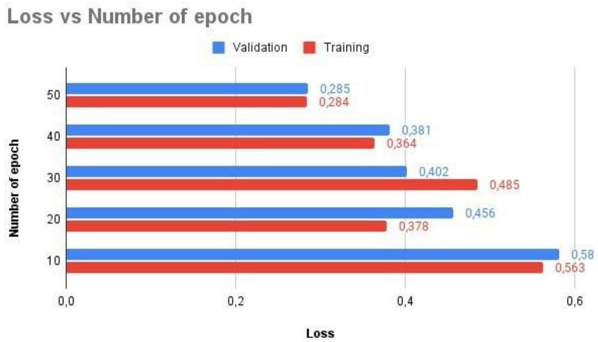


Figure 8: Training and Validation Losses

Figure 9: Example of test images and their predicted values

There are some limitations of the study. Variability in handwriting that means handwriting can vary significantly between individuals, and even within the same individual. This variability can make it challenging to train a model that can accurately recognize all types of handwriting. Also we could notice that training a model can be computationally intensive and requires significant resources, particularly for large-scale applications.

Overall, the findings of the study can be applied in a wide range of real-world scenarios where handwritten text recognition can be used to automate processes, improve accuracy, and reduce processing time, for instance, forms processing, postal services and banking.

In conclusion, it should be noted that the proposed algorithm gives promising results. Modern algorithms that work well in one particular language do not guarantee the same performance in another. Therefore, when building a text recognition model, language specifics should be taken into account.

Abbreviations

AI	Artificial Intelligence
CNN	Convolutional Neural Networks
CTC	Connectionist Temporal Classification
KOHTD	Kazakh Offline Handwritten Text
HMM	Dataset
HTR OCR	Hidden Markov Model
RPN	Handwritten Text Recognition

RRN	Optical Character Recognition Region Proposal Network Recurrent Neural Network

References

1 Elsawy, M. Loey, and H. El-Bakry, "Arabic handwritten characters recognition using convolutional neural network," *WSEAS TRANSACTIONS on COMPUTER RESEARCH*, vol. 5, pp. 11–19, 01 2017.

2 N. Natsagdorj, "Mongolian handwriting character recognition based on convolutional neural network(cnn)." Atlantis Press, 2016/11, pp. 580–584. [Online]. Available: https://doi.org/10.2991/icimm16.2016.105

3 J. R. Prasad, "Handwritten character recognition: A review," pp. 580–584, 2014/10.

4 J. Memon, M. Sami, R. Khan, and M. Uddin, "Handwritten optical character recognition (ocr): A comprehensive systematic literature review (slr)," *IEEE Access*, pp. 1–1, 07 2020.

5 R. Jantayev, S. Kadyrov, and Y. Amirgaliyev, "Complete kazakh handwritten page recognition using start, follow and read method," pp. 580–584, 2021/07. [Online]. Available: https://jatit.org/volumes/Vol99No13/5Vol99No13.pdf

6 S. Kulik, "Neural network model of artificial intelligence for handwriting recognition," pp. 202–211, 2005-2015. [Online]. Available: http://www.jatit.org/volumes/Vol73No2/3Vol73No2.pdf

7 D. Keysers, T. Deselaers, H. A. Rowley, L.-L. Wang, and V. Carbune, "Multi-language online handwriting recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1180–1194, 2017.

8 M. Liwicki, A. Graves, S. Ferna`ndez, H. Bunke, and J. Schmidhuber, "A novel approach to on-line handwriting recognition based on bidirectional long short-term memory networks," sep 2007.

9 J. Pradeep, E. Srinivasan, and S. Himavathi, "Diagonal based feature extraction for handwritten alphabets recognition system using neural network," *International Journal of Computer Science Information Technology*, vol. 3, 02 2011.

10 D. Nurseitov, K. Bostanbekov, M. Kanatov, A. Alimova, A. Abdallah, and G. Abdimanap, "Classification of Handwritten Names of Cities and Handwritten Text Recognition using Various Deep Learning

Models," *Advances in Science, Technology and Engineering Systems Journal*, vol. 5, no. 5, pp. 934–943, 2020.

11 S. Preetha, M. A. I, H. P. Karthik, and S. K. Nishchay, "Machine learning for handwriting recognition," International Journal of Computer (IJC), vol. 38, no. 1, p. 93–101, Jun. 2020. [Online]. Available: https://ijcjournal.org/index.php/InternationalJournalOfComputer/article/view/1637

12 N. Toiganbayeva, M. Kasem, G. Abdimanap, K. Bostanbekov, A. Abdallah, A. Alimova, and D. Nurseitov, "Kohtd: Kazakh offline handwritten text dataset," *Signal Processing: Image Communication*, p. 116827, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0923596522001217

13 M. Kalken, "Handwritten optical character recognition: Implementation for kazakh language."

14 B. Shi, X. Bai, and C. Yao, "An end-to-end trainable neural network for image-based sequence recognition and its application to scene text recognition," *CoRR*, vol. abs/1507.05717, 2015. [Online]. Available: http://arxiv.org/abs/1507.05717

15 D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014. [Online]. Available: https://arxiv.org/abs/1412.6980