Berezovskyi M.Y., Bushyn I.M.
The Bohdan Khmelnytsky National University of Cherkasy

PRE-PROCESSING METHODS EFFECTIVENESS FOR HANDWRITING RECOGNITION BY PHOTO

Abstract. This article presents results of effectiveness comparison of image pre-processing methods for handwriting recognition from an image. It describes recognition accuracy increase/decrease depending on pre-processing methods with the artificial neural network consisting of 5 convolutional layers that use max-pooling between layers and 2 recurrent layers with 256 neurons at the hidden level. The resulting methods combination improves the final accuracy by 7.5% comparing to the initial accuracy.

Keywords: Artificial Neural Network, Convolutional Neural Network, Handwriting Recognition, Image Pre-Processing.

Introduction. Even today's world, there is still a lot of information that exists only in handwriting, although, given the development of information technology, it is much more convenient to analyse, store and transmit information in electronic form. And for this, it is necessary to have reliable mechanisms for converting handwritten text on paper into an electronic version.

The research aims to create an intelligent image pre-processing system for handwriting recognition by analysing image processing methods related to text recognition, their improvement and combination. The task is to analyse the existing methods of image pre-processing for handwriting recognition by photo and to develop intelligent technology by improving and combining existing methods.

When performing the work the following tasks needs to be performed:

• Review and comparative analysis of existing methods of pre-processing images with text.

• Identify the methods that are best for pre-processing images with handwritten text, highlight methods to improve the visibility of the text, divide the photo into individual words and recognition.

• Identify solutions of existing problems in existing methods.

• Design and develop a system for analysis and verification of the implementation of the found methods of image pre-processing for handwriting recognition.

• Conduct experiments using the developed system with selected methods and their combinations, analyse the results of experiments and choose the most optimal mechanisms.

Materials and Methods. Methods such as comparative analysis and experiment were used in the study. Comparative analysis was used to study the feasibility of using certain methods of image pre-processing for handwriting recognition, and the experiment was used to test the effectiveness of different implementations of intelligent pre-processing technology using the analyzed methods and their combinations.

Usually the task of handwriting recognition is not limited to the recognition of one word, which is already perfectly cut from the text, aligned, free of digital noise, etc. The most difficult task is to divide the document into segments that can be recognized by the neural network.

The tasks of preparing a manuscript image for recognition by a neural network can be divided into several groups: general image processing, text alignment, word selection (which in turn, depending on the method, can be divided into line selection and word selection from these lines).

The general processing of the image includes such operations as binarization, reduction of the image to shades of gray, noise reduction, blurring and sharpening of the image, normalization [1], etc.

Results. *Designing the experimental system.*

Experiments are needed to identify the effectiveness of the methods described above, to identify the best of them, and to improve their effectiveness. But given that experiments require the training of a neural network of the same configuration with different sets of pre-processing methods, it is first necessary to develop a system that will allow these experiments to be performed.

Therefore, a neural network model was developed to test the effectiveness of image pre-treatment methods for handwriting neural network training. [2]

It consists of 5 layers of CNM, 2 layers of LNM, CTC level loss, and decoding.

• The input image is gray and has a size of 128x32 [3]

• 5 layers of ZNM display the input image to a sequence of objects of size 32x256

• 2 layers of RNM with 256 units distribute information through the sequence and display the sequence to a matrix of 32x80. Each matrix element represents an estimate for one of the 80 symbols at one of the 32-time steps

• The CTC layer calculates the value of the losses specified by the matrix and the true text (during training) or decodes the matrix to the final text with the decoding of the best path. [6]

Implementation depends only on the numpy, cv2, and tensorflow libraries. The experiments will be carried out taking into account the following points:

• The batch size is set to 50

• One epoch - 500 games, so one epoch will include 25,000 images

• IAM Handwriting Database [5] with 118 thousand images will be used for neural network training

• For training and validation, the dataset is divided into 95% for training and 5% for validation. Data for validation remain unchanged, 25 thousand images are selected from 95% randomly

• 2 parameters will be evaluated: the average error in word recognition and the error in the characters in the recognized words

• If the recognition accuracy does not improve after 10 epochs, the training ends

• Accuracy will be considered the average error of character recognition.

Base model

First you need to determine the starting point from which you can start. To do this, we will train the neural network, the pre-processing of the image for which will include only the conversion of the image into shades of gray.

In order to make sure that the results are not random and based on regularity, training will be conducted several times to determine how the accuracy of accuracy results when learning from scratch with the same data set and the same method of pre-processing.

The learning results of the neural network are presented in table 1 and in Fig. 1.

Table 2.1

	Base Model - 1	Base Model - 2
Word error	28,80%	28,94%
Symbol error	11,96%	11,95%

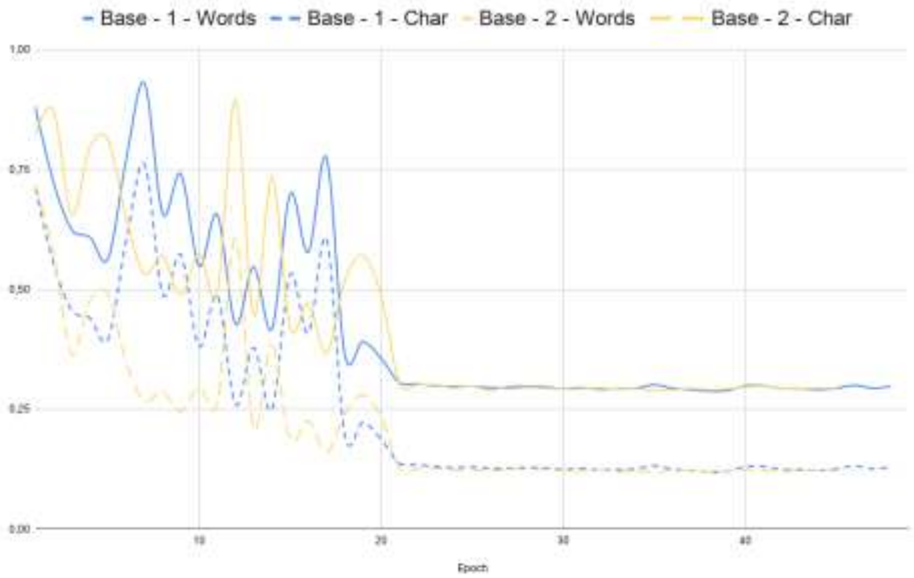

Fig.1. Diagram of the average error of the neural network depending on the era for the base model

As can be seen from the results, the accuracy of the neural network hardly changes when re-learning with random initial weights in the model, but with the same set of data for training and the same method of pre-processing.

The first approximately 20-25 epochs of learning are unstable with constant fluctuations in accuracy, but after 25 epochs only slightly improves accuracy. It is worth noting this feature of the basic model and compare the methods used in the following experiments.

The difference in results between the two re-training of NN is only 0.01%, which can be considered as an error in future experiments. Also, any change greater than 0.1% in character recognition accuracy can be considered as an effect of the pre-processing method.

For the model for comparison the second run of training will be used as accuracy it turned out a little higher.

Final model

After experiments, it was found that binarization not only does not improve the result in the case of handwriting recognition, but also worsens the results of handwriting recognition from the image.

At normalization it is possible to see obvious improvement, both on stability of results, and on the maximum accuracy. Although the absolute value of accuracy has not changed much, but the stability of the result has significantly improved. This improvement will reduce training time when using very large amounts of data, as it allows you to stop learning before the maximum accuracy has been reached (and determining such a maximum for a large sample is also a difficult task).

When using augmentation, the absolute accuracy result improved markedly relative to the base model. The change in the relatively normalized model is not so noticeable, but it is also there. But the other side of improving accuracy was the much lower stability of the results.

Taking into account the previous results, only normalization and swelling of the images were chosen for the combination. Normalization significantly improves the stability of learning and at the same time increases the accuracy of recognition. When the images are inflated, an increase in accuracy can be achieved, but the stability of the results deteriorates even compared to the base model.

The results in comparison with the base model are presented in Table 2 and in Fig. 2.

Table 2

	Base Model	Normalization + Augmentation
Word error	28,94%	26,92%
Symbol error	11,95%	10,88%

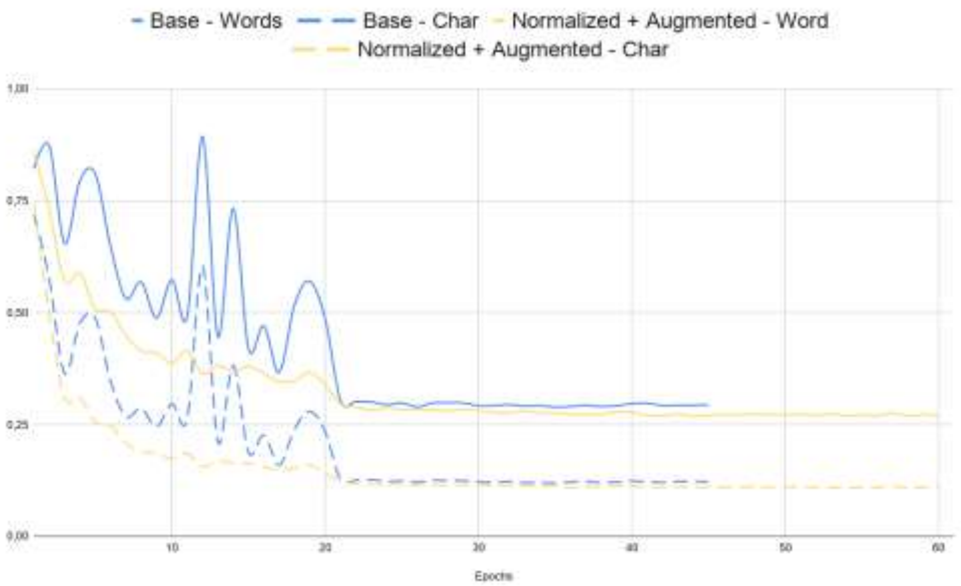

Fig.2. Diagram of the average error for the base image and using inflation

The combination of the methods described above is expected to give results, both in a marked increase in accuracy and in improving the stability of learning relative to simply inflated images.

The final result showed an improvement of 7.5% in word recognition accuracy and 9.83% in character recognition accuracy relative to the base model.

Summary & Conclusions.

The article carried out a theoretical study of image pre-processing methods for handwriting recognition and formed a set of methods, the effectiveness of which was analysed at the stage of experiments.

An experiment was conducted to analyse the effectiveness of various methods of image pre-processing to teach a neural network for handwriting recognition. As a result of the experiment, the result relative to the basic model trained in grayscale images without additional processing was able to improve the result by 7.5% in word recognition accuracy and 9.83% in character recognition accuracy, which can be considered a significant result.

As a result, the following prerequisites are formed for handwriting recognition systems:

• For neural network training, binarization leads to data loss (which does not preclude the use of binarization in the early stages of processing)

• Normalization and bloating can significantly increase the accuracy of recognition.

Discussion. This approach has further prospects for improving accuracy. This can be achieved by sharpening images, whitening the background, increasing saturation, deslanting, etc. This will allow the algorithm to be used to digitize library archives automatically. Also, using this algorithm, you can optimize processes in schools by automating handwriting verification.

References

1. Rafael C. González, Richard Eugene Woods (2007). Digital Image Processing. Prentice Hall. p. 85. ISBN 978-0-13-168728-8.
2. Yun Wang, Juncheng Li, Florian Metze. Comparing the Max and Noisy-Or Pooling Functions in Multiple Instance Learning for Weakly Supervised Sequence Learning Tasks. April 3, 2018. pp. 5 URL: https://arxiv.org/pdf/1804.01146.pdf.
3. Blumenstein, Michael, Brijesh Verma, and Hasan Basli. "A novel feature extraction technique for the recognition of segmented handwritten characters." In Document Analysis and Recognition, 2003. Proceedings. Seventh International Conference on, pp. 137-141. IEEE, 2003
4. Eikvil, Line. "Optical character recognition." citeseer. ist. psu. Edu / 142042. Html (1993).
5. IAM Handwriting Database," IAM Handwriting Database - Computer Vision and Artificial Intelligence. [Online]. Available: http://www.fki.inf.unibe.ch/databases/iam-handwriting-database. [Accessed: 18-Sep-2019].
6. GuifangLina, WeiShen. Research on convolutional neural network based on improved Relu piecewise activation function. Volume 131, 2018, Pages 977-984 URL: https://www.sciencedirect.com/science/article/pii/S1877050918306197.
7. Casey, Richard G., and Eric Lecolinet. "A survey of methods and strategies in character segmentation." Pattern Analysis and Machine Intelligence, IEEE Transactions on 18, no. 7 (1996): 690-706.

Bushyn I. M.

The Bohdan Khmelnytsky National University of Cherkasy

Анотація. У цій статті представлені результати порівняння ефективності методів попередньої обробки зображень для розпізнавання рукописного вводу із зображення. Він описує збільшення / зменшення точності розпізнавання залежно від методів попередньої обробки за допомогою штучної нейронної мережі, що складається з 5 згорткових шарів, які використовують макс-пул між шарами та 2 повторюваних шари з 256 нейронами на прихованому рівні. Отримана комбінація методів покращує кінцеву точність на 7,5% порівняно з початковою точністю.

Ключові слова: розпізнавання рукописного тексту, штучна нейронна мережа, згорткова нейронна мережа, попередня обробка зображень.