Arabic Manuscript Author Verification Using Deep Convolutional Networks

Andrei Boiarov[1], Alexander Senov[1] and Alexander Knysh[23]
[1] Faculty of Mathematics and Mechanics, Saint Petersburg State University, Saint Petersburg, Russia
e-mail: andrei.boiarov@gmail.com, alexander.senov@gmail.com
[2] Department of Near Eastern Studies, University of Michigan, Ann Arbor, Michigan, USA
[3] Research Laboratory for Analysis and Modeling of Social Processes, St. Petersburg State University, St. Petersburg, Russia
e-mail: alknysh@umich.edu

***Abstract*— In this paper, we propose an automatic method for manuscript author verification based on an analysis of consecutive patches extracted from an image. The classification algorithm uses a deep convolutional network with two types of patch extraction: one based on connected components and the other based on usage of a fixed-size sliding window. We apply this method to verify the authorship of the Arabic manuscript entitled *al-Khitat* attributed to the hand of the renowned medieval Arab historian al-Maqrizi. Using appropriately collected ground-truth labeled data for convolutional network training purpose our method has demonstrated promising results when applied to previously unseen manuscripts.**

***Keywords*— Author verification, Convolutional networks, Deep learning, Handwriting recognition, Historical Arabic manuscript.**

I. Introduction

The problem of manuscript author verification is quite important nowadays. Since manual verification has a subjectivity drawback, the solution requires objective automatic methods. The present study was motivated by the recent discovery by Dr. Noah Gardiner of the holograph (autograph) copy of the third volume of al-Maqrizi's famous "Description of Egypt" in the Library of the University of Michigan [1]. The full title of the manuscript is "The Book of Admonitions and Lessons in the Catalogue of Territorial Divisions and Historical Monuments"~~;~~ usually cited as *al-Khitat*. The manuscript was finished shortly after 831 A.H. (1427 C.E.) by the celebrated Egyptian historian Taqi al-Din Ahmad Ibn ’Ali al-Maqrizi (d. 845/1442). In his paper, Dr. Gardiner made the assumption that it might be a draft copy of the work. He then sent images of the codex to Prof. Frederic Bauden of the University of Liege, the author of numerous articles on al-Maqrizi autographs. B~~ua~~den identified it as the fair copy (the author's final version) of the third volume of *al-Khitat*, and thus the only fair copy of any volume of *al-Khitat* to have been found.

Given the importance of this discovery for the history of science (al-Maqrizi's *al-Khitat* is one of the earliest descriptions of the topography of Cairo and ancient Egyptian monuments in its environs as well as Alexandria), authors decided to verify Dr. Gardiner's and Prof. Bauden's findings using method based on deep convolutional networks.

Previously used methods for author verification of Arabic manuscript and for related fields were focused on developing various types of features that can be obtained from a manuscript picture. These features are then used to learn a classification model. Here are some of the approaches: textural, allographic features and clustering methods [2], [3]; contour-based, oriented basic image, K-SIFT features and SVM [4], texture-related and structure-related features [5], TF/IDF and clustering algorithms [6]. In this paper, we present a novel approach based on learning of deep hierarchical structure of features from the row image. This method belongs to the class of the deep learning algorithms [7] and uses convolutional network [8] for feature extraction and prediction. Deep convolution networks since 2012 year [9] became the state of the art in many areas of computer vision: objects recognition, face identification, optical character recognition, object detection, etc [7]. Deep learning approach is based on using artificial neural networks with a ~~huge~~ number of layers which allows to train deep hierarchical representation of data. This representation in many cases works better then handmade features [7], [9], [10]. Moreover deep networks are not ~~saturate~~ with an increasing amounts of data thus they have better generalization ability then traditional "shalow" methods like logistic regeression or SVM [7].

The paper is organized as follows. In Section II, a description of the data set used in experiments is given. Section III contains presentation of our methodology of patches from image extraction, training deep convolution networks and making decision of manuscript author. In Section IV, we present the results of our experiments.

II. Data Set

Solving the problem of ~~al-Maqrizi~~ authorship ~~verification~~ requires training and test data sets. We consider a single page of the manuscript as one data unit. As a one unique data element we consider single page of manuscript. The data set consists of two sets of manuscript materials.

- Two holographs ~~of~~ al-Maqrizi verified by Prof. Bauden will serve as a positive example.
- Eight manuscripts not written by al-Maqrizi's hand (taken from the University of Michigan Hatcher Library Special Collections) are considered to be a negative example. These manuscripts were selected in such a way that the date and place of writing of each of them would be close

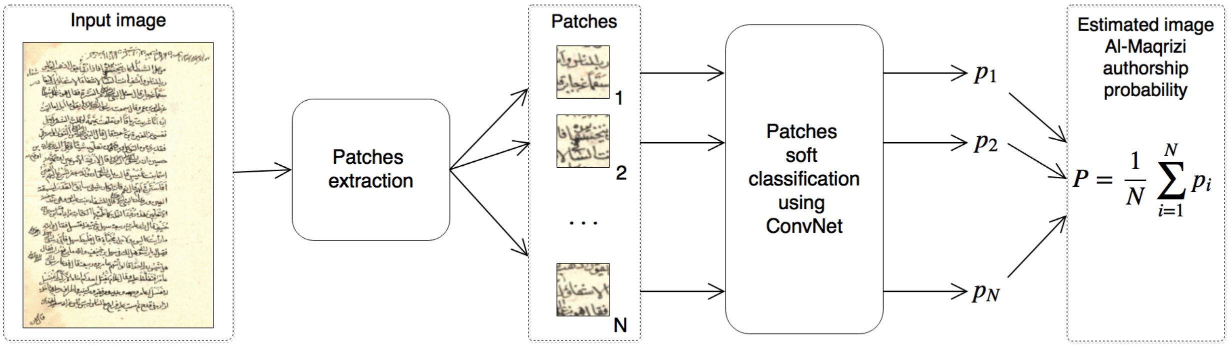

Fig. 1. al-Maqrizi authorship classification pipeline

to those of al-Maqrizi's *al-Khitat*, namely, the 14th and 15th centuries, Egypt and Syria.

To enhance robustness of the learning process the data obtained were divided into training and test sets:

- Training set: 1 al-Maqrizi's manuscript consisting of 26 pages and 5 ~~not al Maqrizi's~~ manuscripts each of which consists of 7 pages.
- Test set: 1 al-Maqrizi's manuscript consisting of 14 pages and 3 non-al-Maqrizi's manuscripts each of which consists of 7 pages. Authors of these 3 manuscripts differ from authors of 5 negative examples in the training set.

Two examples from the training set is illustrated ~~at~~ Figure 2.

It is important to note that we split our data set by the factor of the manuscript author. When the algorithm is learning on the training set, it ~~can not~~ see the authors in the test set. Thus, we enforce our method to infer general properties al-Maqrizi's handwriting not specific ~~for~~ particular manuscript. ~~Number~~ of non-al-Maqrizi's documents is chosen in such way that positive and negative classes in training and test sets are roughly balanced.

The goal of this study is to verify the authorship of the *al-Khitat* manuscript that consists of 32 pages.

III. Method

We consider author verification as a binary classification problem: positive class designated as 1 ~~consist~~ of images that belong~~s~~ to al-Maqrizi's hand and negative class designated as 0 consist of images of different author. In this context our goal is to build a classification pipeline able to estimate the probability (*soft* classification) that a given image belongs to the 1 (al-Maqrizi) class. Thus, perfect classifier should return 1 for all al-Maqrizi's images and 0 for all non-al-Maqrizi's images. The entire proposed classification pipeline for al-Maqrizi authorship verification illustrated ~~at~~ Figure 1 ~~and~~ consists of the following steps:

1) Image preprocessing.
2) Extracting patches from candidate image.
3) Soft classification of patches using convolution network.
4) Averaging predicted patches probabilities to produce overall candidate image of the probability of al-Maqrizi's authorship.

Fig. 2. Examples of al-Maqrizi (left) and not al-Maqrizi (manuscript from Cairo, 1466 C.E.) (right) holographs from training set

Each of these steps are ~~thoroughly~~ described in the following sections.

A. Image preprocessing

Image preprocessing was done to bring all images to relatively same size and scale by performing the following steps:

1) Removing part of image within the text bounding box.
2) Resizing resulted image to fixed resolution (700×500).

~~This~~ steps are reasonable enough since all images from our data set have approximately equal number of text lines, relative font size and text bounding box aspect ratio. Thus, removing blank margin from image and resizing it to fixed size we make our classification problem ~~more simple~~ since images became less diverse.

Note: since this step is done only to unify our images data set we did not include it ~~to~~ the pipeline ~~at~~ Figure 1.

B. Patches extraction

Patches extraction is a process of breaking given image into set of potentially overlapping sub-images called patches. The basic idea is that a patch should represent a small but yet meaningful part of image for the purpose of authorship verification. On the one hand patches should be as small as possible to reduce dimension of input vector for classification, but on the other hand patches still should contain enough information for successful authorship verification.

We use two alternative methods for patches extraction. Both of them described below.

1) Sliding window based method: This method ~~use~~ sliding window of the size 80×80 pixels, a stride of size 20 pixels and operates in the following way:

0) Sliding window moved to top left conner of the image
1) Patch extracted from current window position
2) Sliding window shifts to the right on the stride size, if possible. Otherwise it moved to the ~~most left~~ position and shifted down on the stride size.
3) Goto 1.

Figure 3 demonstrates and example of ~~resulted~~ patch.

2) Connected components-based method: This method attempts to extract patches containing one or~~e~~ few symbols on the same text line by utilizing idea of connected components. Its general scheme described below.

1) First, input image binarized using Otsu's filter [11]. After this step images consist of only 1's (symbols) and 0's (blank space).
2) Then, connected components of 1's extracted from a binarized image using the algorithm from [12].
3) ~~Too small, too big and too stretched connected components~~ filtered out using several empirical rules, e.g.: major axis to minor axis ratio less then 10, minimum minor axis length greater then 3 pixels, maximum major axis length less then 220 pixels and minimum major axis length greater then 5.
4) Points of connected ~~components~~ centers clustered using DBSCAN clustering algorithm [13], [14] with following parameters: epsilon = 0.5 and minimum number of points to form a dense cluster = 10. ~~Then~~ connected components corresponding to the points which ~~did~~ not fall in any cluster are labelled as outliers and filtered out.
5) Finally, remaining connected components bounding boxes extracted from the source image and resized to 28×28 pixels size.

Example of connected components based patches shown on Figure 4.

As one can see, connected components-based patches usually consist of one or ~~few~~ letters thus providing a high robustness for different image scale and size in contrast to fixed-size sliding window patches (since all images been preprocessed this feature~~s~~ is not important in our case). However, sliding-window patches contain much more information: several symbols from multiple lines — a very important feature for authorship verification.

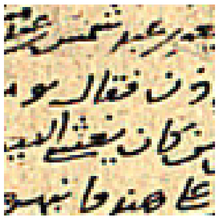

Fig. 3. Sliding window patch example

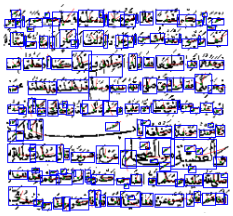

Fig. 4. Connected components patches example

C. Deep convolution network

Denote x_i as the i-th patch obtained from an image in the previous step and y_i as binary label. Then for predicting the probability

$$p_i = P(y_i|x_i)$$

that the current patch belongs to al-Maqrizi's hand we need to train a discriminative model. As a robust and powerful classification algorithm deep convolution networks [7], [8] are chosen.

In a case of sliding window patches we apply AlexNet type of network [9]. Its architecture and layers parameters is described in Table 1. This table contains the following columns: layer type, patch size and stride parameters for convolutional and pooling layers, number of layer output neurons. Network architecture is illustrated on Figure 5.

As an input network takes 80×80 RGB image. All activation functions in convolution and fully connected layers are rectified linear units (ReLU): $f(x) = \max(0, x)$. This deep network ~~was~~ trained by stochastic gradient descent method with Nesterov momentum, initial learning rate is 0.01 and ~~it~~ decrease policy is step down.

Table 1. Sliding window patch convolution network

layer type	**patch size/ stride**	**output number**
convolution	11 / 4	96
local response norm		96
max pool	3 / 2	96
convolution	5 / 1	256
local response norm		256
max pool	3 / 2	256
convolution	3 / 1	384
convolution	3 / 1	384
convolution	3 / 1	256
max pool	3 / 2	256
fully connected		4096
dropout (50 %)		4096
fully connected		4096
dropout (50 %)		4096
fully connected		2
softmax		2

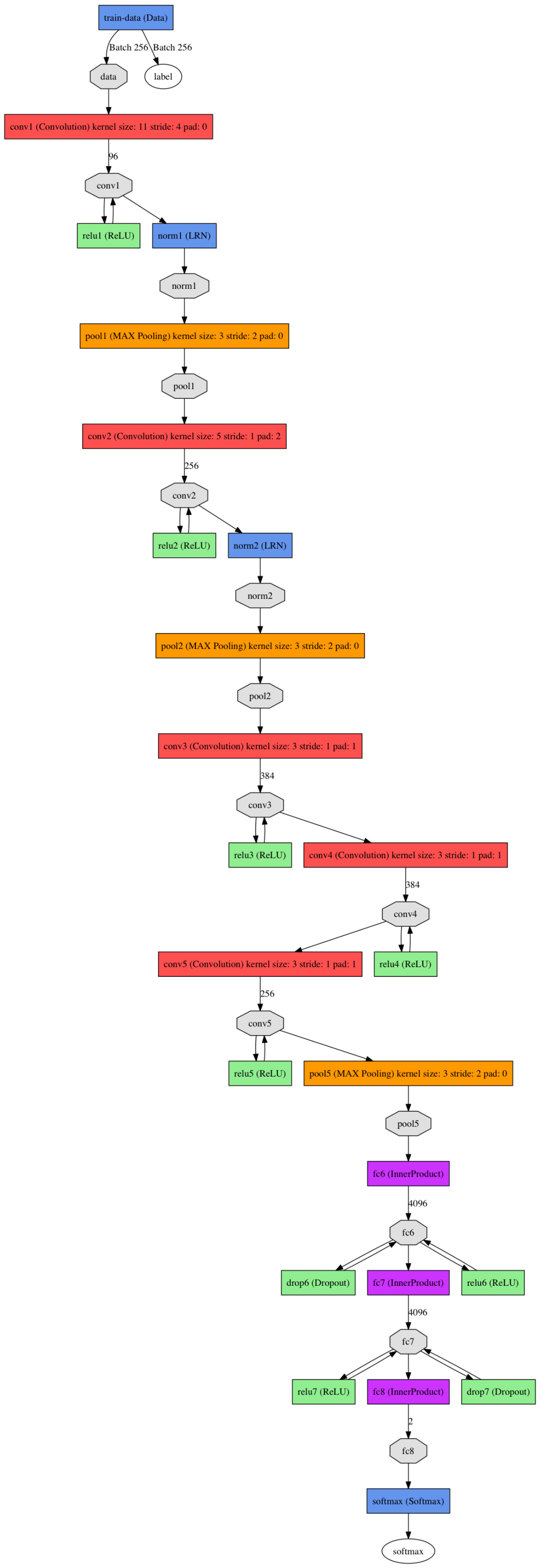

Fig. 5. Sliding window patch convolution network architecture

Table 2. Connected components patch convolution network

layer type	patch size/ stride	output number
convolution	7 / 2	64
max pool	3 / 2	64
local response norm		64
convolution	1 / 1	64
convolution	3 / 1	192
local response norm		192
max pool	3 / 2	192
inception		256
inception		480
max pool	3 / 2	480
inception		512
convolution	1 / 1	128
fully connected		1024
dropout (70 %)		1024
fully connected		2
softmax		2

With connected components patches we use GoogLeNet type of deep convolutional network [10]. Its architecture and layers parameters ~~is provided~~ in Table 2.

As an input network takes 28 × 28 RGB image. All activation functions in convolution and fully connected layers are rectified linear units (ReLU). Inception layer is a combination of several convolution and pooling layers, ~~it~~ is pictured on Figure 6; for more details see [10]. Learning method for this convolutional network is stochastic gradient descent with Nesterov momentum and initial learning rate 0.01 with step down decrease policy.

IV. Results and Discussion

We have experimented with both types of patches extraction and corresponding convolutional networks.

A. Deep convolutional network learning

Deep artificial networks have been trained with backpropagation algorithm [8] and shown small training error. For sliding window patch convolution network accuracy value on the test set is near 87%. For connected components patch convolution network accuracy is worse — 80%.

During training process deep convolutional network is trying to build a hierarchical structure of data presentation. On the first layers trained low-level features, on last layers – some high-level representation of an input data. This approach allows to build classifiers and successfully meet ~~the~~ challenges supervised learning. To illustrate this idea, consider features activation of the first local response normalization layer (LRN) in the sliding window patch convolution network after feed forward through the network sliding window patch from Figure 3. As can be seen in Figure 7 some feature maps are spiking to various arabic words or combinations of words. This behavior demonstrates the purpose of this layer in the structure of the entire deep network.

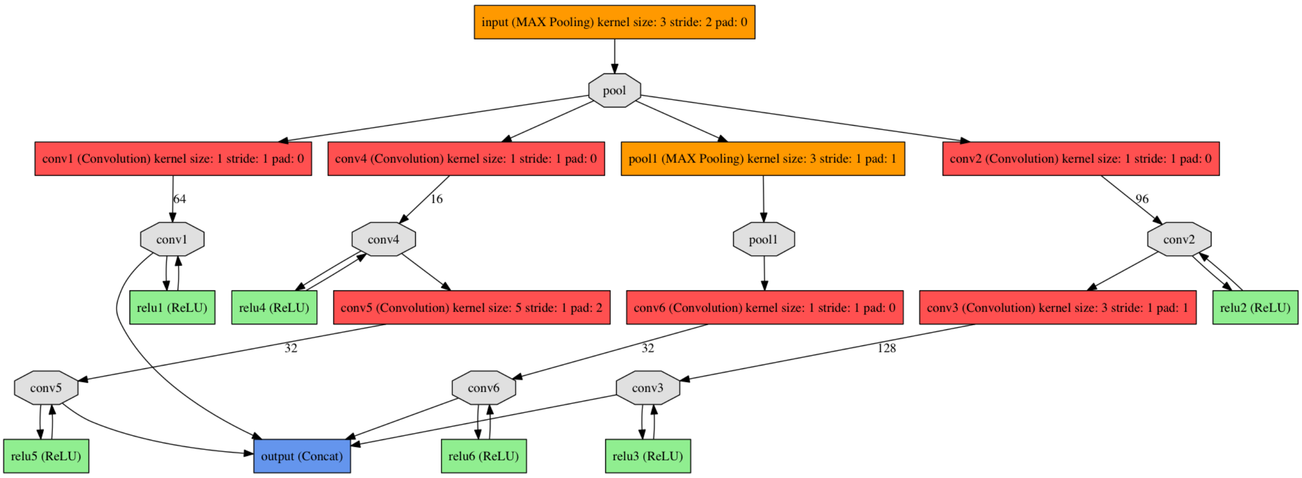

Fig. 6. Inception layer architecture

B. Manuscript classification

To assess quality of the classification pipeline we have used only images from the test set, since they are the only ones not used in the learning process.

Figure 8 demonstrates the classification results for two images from the test set: one from al-Maqrizi class and one from non-al-Maqrizi class. As one can see, both of them can be classified quite confidently as ~~having~~ al-Maqrizi ~~as~~ the author with the probability 0.0085 and 0.94 respectively. Figure 9 shows the classification result for an *al-Khitat* manuscript page. It gives 0.77 probability of al-Maqrizi being the author.

Regarding the entire test set classification for sliding window patches with decision threshold 0.5 we obtained precision 0.99 and recall 0.92. The method based on connected components patches is less robust: it generates many false positive predictions. This approach has a great potential, however,

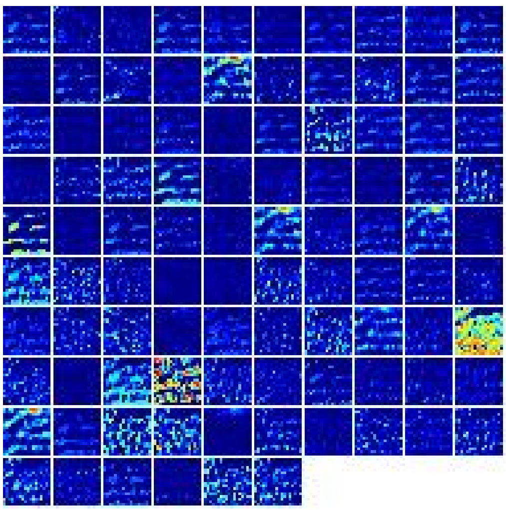

Fig. 7. Features activations on the first local response normalization layer (LRN) in the sliding window patch convolution network

Fig. 8. Sliding window patches al-Maqrizi authorship classification example visualization on non-al-Maqrizi manuscript page (left) and al-Maqrizi manuscript page (right). Patches probabilities are visualized by using white-red colors (corresponding to 0-1 classes) on the top of the original image. Average al-Maqrizi authorship probability across patches is 0.0085 and 0.94 respectively.

its improvement requires many more training examples. In sum, convolutional network ~~learned~~ on connected components patches is a promising field for further research.

By using sliding window patches with deep network method we have arrived at the conclusion that the mean probability of al-Maqrizi's authorship ~~of~~ the entire set of the *al-Khitat* pages equals 0.86.

V. Conclusion

~~Deep~~ convolutional network with sliding window patches demonstrates very prominent result in resolving the problem of the authorship of *al-Khitat*. In comparison with other methods [2], [3], [4], [5] our approach shows a positive outcome without the need to construct complex features. To apply this method to other Arabic ~~authors,~~ we will train our network

Fig. 9. Sliding window patches al-Maqrizi authorship classification example visualization on the page from *al-Khitat* manuscript. ~~Average~~ al-Maqrizi authorship probability across patches is 0.77.

on a larger volume of data. The joint study of al-Maqrizi's "Description of Egypt" undertaken by a historian-philologist and mathematicians is a unique experiment in working across disciplinary boundaries to achieve a common goal. The results obtained thus far bode well for the future by opening new horizons for scholars of "Oriental" manuscripts who often desperately lack resources (other than their own eyes and intuition) to verify the provenance and authorship of the manuscript material they are working with.

ACKNOWLEDGMENT

This research is supported by Saint Petersburg State University grant 6.37.181.2014. The authors express their deep gratitude to Mrs. Evyn Kropf of the Hatcher Graduate Library who kindly facilitated access to the University of Michigan library resources.

REFERENCES

[1] N. Gardiner, F. Bauden, "A Recently Discovered Holograph Fair Copy of al-Maqrizi's al-Mawa'iz wa-al-i'tibar fi dhikr al-khitat wa-al-athar (Michigan Islamic MS 605)," in *Journal of Islamic Manuscripts*, vol. 2, E. J. Brill, Leiden and Boston, 2011, pp. 123–131.

[2] M. Bulacu, L. Schomaker, A. Brink, "Text-independent writer identification and verification on offline arabic handwriting," in *Proc. 9th International Conference on Document Analysis and Recognition, ICDAR*, Curitiba, 2007, pp. 769–773.

[3] M. Bulacu, L. Schomaker, "Automatic Handwriting Identification on Medieval Documents," in *Proc. of the International Conference on Image Analysis and Processing*, Modena, 2007, pp. 279–284.

[4] D. Fecker, A. Asi, W. Pantke, V. Margner, J. El-Sana, T. Fingscheidt, "Document Writer Analysis with Rejection for Historical Arabic Manuscripts," in *Proc. 14th nternational Conference on Frontiers in Handwriting Recognition, ICFHR*, Crete, 2014, pp. 743–748.

[5] S. Godoy-Calderon, E. M. Felipe-Riveron, E. C. Herrera-Luna, "Multilevel Modeling of Manuscripts for Authorship Identification with Collective Decision Systems," *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*, vol. 7441, pp. 757–764, 2012.

[6] E. Dunn, "Computational Methods for Determining the Similarity between Ancient Greek Manuscripts,"Master of Science thesis, Dept. Computer Science., Dept. Inform. Systems and Operations Manag. Univ. of North Carolina, Wilmington, 2012.

[7] Y. Lecun, Y. Bengio, G. Hinton, "Deep learning," *Nature*, no. 521, pp. 436–444, May. 2015.

[8] Y. Lecun, L. Bottou, Y. Bengio, P. Haffner, "Gradient-based learning applied to document recognition," in *Proc. of the IEEE*, 1998, pp. 2278–2324.

[9] A. Krizhevsky,I. Sutskever, G. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," in *Advances in Neural Information Processing Systems*, vol. 25, 2012, pp. 1097–1105.

[10] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, "Going deeper with convolutions," in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, Boston, 2015, pp. 1–9.

[11] N. Otsu, "A threshold selection method from gray-level histograms, " *Automatica*, vol. 11, 1975, pp. 285–296.

[12] C. Fiorio, J. Gustedt, "Two linear time union-find strategies for image processing, " *Theoretical Computer Science*, vol. 154, no. 2, 1996, pp. 165–181.

[13] M. Ester, H.P. Kriegel, J. Sander, X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise, " in *Proc. of Second International Conference on Knowledge Discovery and Data Mining*, vol. 96, no. 34, 1996, pp. 226–231.

[14] O. Granichin, V. Volkovich, D. Toledano-Kitai, *Randomized Algorithms in Automatic Control and Data Mining*. Springer-Verlag: Heidelberg, New York, Dordrecht, London, 2015, 251 p.