Arabic Manuscript Author Verification Using Deep Convolutional Networks

Andrei Boiarov*, Alexander Senov†, Alexander Knysh‡ and Dmitry Shalymov§
*†§ Faculty of Mathematics and Mechanics
Saint Petersburg State University
Saint Petersburg, Russia
Email: *a.boiarov@spbu.ru, †alexander.senov@gmail.com, §dmitry.shalymov@gmail.com
‡Department of Near Eastern Studies
University of Michigan
St. Petersburg State University Laboratory of Analysis and Modeling of Social Processes
Ann Arbor, Michigan, USA
Email: alknysh@umich.edu

***Abstract*—The problem of manuscript author verification is quite important nowadays and since manual verification has a subjectivity drawback, need in objective automatic methods arises. In this paper we propose an automatic method for author verification based on consecutive patches from image extraction and such state of the art image classification algorithm as deep convolutional network with two types of patches extraction: connected components based and fixed-size sliding window based. We apply this method to specific problem of medieval Egyptian arabic historian al-Maqrizi's authorship verification of manuscript al-Khitat which volume was recently discovered. Using appropriately collected ground-truth labelled data for convolutional network training purpose our method demonstrates very promising results on previously unseen manuscripts.**

I. Introduction

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 (Michigan Islamic MS 605) [1]. The full title of the manuscript is "al-Mawa'iz wa-al-i'tibar fi dhikr al-khitat wa-al-athar" ("The Book of Admonitions and Lessons in the Catalogue of Territorial Divisions and Historical Monuments"; usually cited simply as al-Khitat). The manuscript was copied well after 818 A.H. (1415 C.E.) and was finished shortly after 831 (1427) by the celebrated Egyptian historian Taqi al-Din Ahmad Ibn 'Ali al-Maqrizi (d. 845/1442). It is the only known Maqrizi holograph (autograph) in the Americas. A number of elements in the codex, including the apparent age of the paper, lacunae in the text where the dates of certain events had not been filled in, and a number of marginal addenda and sewn-in inserts containing text found in the printed editions, led Dr. Gardiner to suspect that it might be a draft copy of the work. He visually collated the predominant hand of the codex and the inserts with some published images of al-Maqrizis hand and felt that a match was highly likely. 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. Prof. Bauden confirmed that the codex was indeed copied by al-Maqrizi himself, and was thus a holograph (autograph). He identified it as the fair copy (the authors 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 Khitat is one of the earliest descriptions of the topography of Cairo and ancient Egyptian monuments in its environs as well as Alexandria), a cross-disciplinary team of researchers affiliated with the St. Petersburg State University (Laboratory of Analysis and Modeling of Social Processes) decided to verify Dr. Gardiners and Prof. Baudens findings by using method based on deep convolutional networks.

Previously used methods for author verification of arabic manuscript and for related fields were focused on developing of various types of features that can be obtained from a manuscript picture [2], [3]. 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 deep learning algorithms [4] and uses convolutional network [5] for feature extraction and prediction. Deep convolution networks since 2012 year [6] became the state of the art in many areas of computer vision: objects recognition, face identification, optical character recognition, object detection etc [4].

The paper is organized as follows. In section II a description of 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 our experiments results are presented. Finally, conclusion is given.

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.

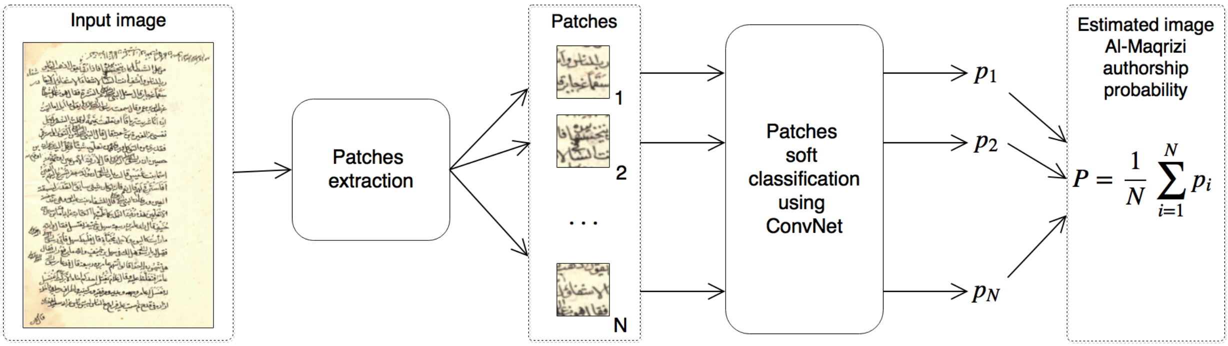

Figure 1: al-Maqrizi authorship classification pipeline

II. Data Set

Solving of the problem of al-Maqrizi authorship verification requires training and test data sets. As a one unique data element we consider single page of manuscript. Data set composes of a set of consistent parts from manuscripts of two kinds:

- Two sets of verified al-Maqrizi's holographs from codex of Prof. Bauden is taken as a positive examples.
- Eight manuscripts not written by al-Maqrizi's hand from the University of Michigan Hatcher Library (Special Collections) are considered as a negative examples. This manuscripts are selected in such way that the date and place of writing each of them is close to al-Maqrizi's Khitat: 14th and 15th centuries, Egypt and Syria.

For robustness of learning process obtained data is 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 not al-Maqrizi's manuscripts each of which consists of 7 pages. Authors of this 3 manuscripts differ from authors of 5 negative examples in training set.

It is important to note that we split our data set by the factor of a manuscript author. Thus when algorithm is learning on training set, it can not see authors in test set. In that way we achieve robustness our method in terms of manuscripts because input of method is a whole manuscript. Number of not al-Maqrizi's documents is chosen in such way that training and test set is nearly balanced.

Main interest of this paper is author verification of al-Khitat manuscript consisting of 32 pages.

III. Method

We consider author verification as a binary classification problem: al-Maqrizi class denoted as 1 and non al-Maqrizi class denoted as 0. In this context our goal is to build a classification pipeline able predict probability (*soft* classification) that given image belongs to the 1 (al-Maqrizi) class. The entire al-Maqrizi authorship classification pipeline illustrated at figure 1 consists of the following steps:

0) Image preprocessing.
1) Extracting patches from candidate image.
2) Patches soft classification using convolution network.
3) Averaging predicted patches probabilities to produce overall candidate image al-Maqrizi authorship probability.

Each of this steps are thoroughly described in the following sections.

A. *Image preprocessing*

This step was done to bring all images to relatively same size and scale by performing following steps:

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

This steps are reasonable enough since all images from our data set has approximately equal number of text lines and text bounding box aspect ratio.

Since this step is done only to unify our images data set we does not include it in the pipeline.

B. *Patches extraction*

The patches extraction method generates a set of sub-images called patches from given image. The basic idea is that patch should represent small but yet meaningful part of image for the main purpose — authorship verification. We use two alternative methods for patches extraction described in following subsections.

1) Sliding window based method: This method splits image into patches by a grid of fixed cell of size 80×80 pixels with a stride 20 pixels. Figure 2 illustrates the idea.

2) Connected components based method: This method uses following routine for patches extraction

1) Input image binarization using Otsu's filter [8].
2) Connected components extraction from binarized image using algorithm from [9].
3) Too small, too big and too stretched connected components filtering using several empirical rules, e.g.: major axis to minor axis ratio less then 10, minimum minor axis length greater then 3 pixels, etc.
4) Outlier connected components filtering using DBSCAN clustering algorithm [10], [11].

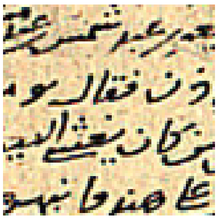

Figure 2: Sliding window patch example

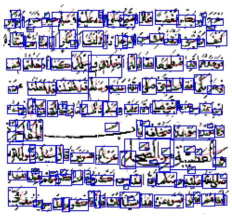

Figure 3: Connected components patches example

5) Extracting remaining connected components bounding boxes from source image and resizing them 28×28 pixels size.

Example of connected components based patches show on figure 3.

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

C. Deep convolution network

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

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

that 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 [4], [5] are chosen.

In a case of sliding window patches we apply Alexnet type of network [6]. Architecture of this network described in table I.

Table I: Sliding window patch convolution network

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

As an input network takes 80×80 RGB image. All activation functions in convolution and fully connected layers is rectified linear units (ReLU). 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.

With connected components patches we use GoogLeNet type of deep convolutional network [7]. It architecture provided in table II.

Table II: Connected components patch convolution network

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

As an input network takes 28×28 RGB image. All activation functions in convolution and fully connected layers is rectified linear units (ReLU). Inception layer is a combination of several convolution and pooling layers, for more details see [7]. 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 experimented with both types of patches extraction and corresponding convolutional networks.

A. Deep convolutional network learning

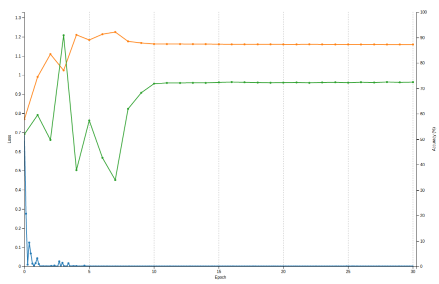

Figure 4: Sliding window patch convolution network learning process.

Figure 4 demonstrates values of loss function on training (blue curve) and test (green curve) sets depending on the learning epoch for sliding window patch convolution network. Also one can see accuracy (orange curve) on the test set. It final value is near 87%. For connected components patch convolution network accuracy is worse — 80%.

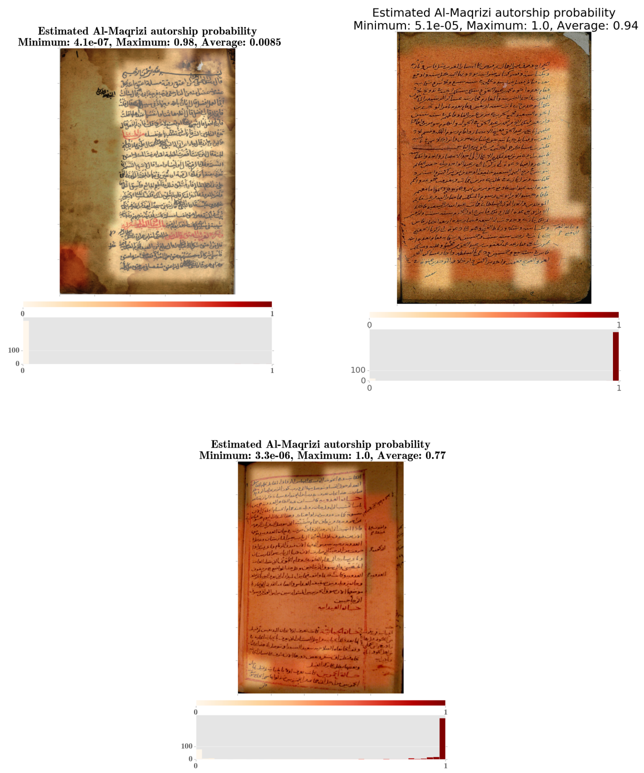

Figure 5: Sliding window patches al-Maqrizi authorship classification example: not al-Maqrizi manuscript page (left), al-Maqrizi manuscript page (right) and manuscript page from Khitat (bottom). Patches probabilities visualized using white-red colors (corresponding to 0-1 classes) on top of the original image. Additionally, probabilities histogram presented below.

B. *Manuscript classification*

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

Figure 5 demonstrates classification result for two images from the test set: one from al-Maqrizi class and one from not al-Maqrizi class. As you can see, both of them classified quite confidently with estimated al-Maqrizi authorship probability 0.0085 and 0.94 respectively. Also this figure shows classification result for al-Khitat manuscript page. It gives 0.77 probability of al-Maqrizi authorship.

Regarding entire test set classification for sliding window patches with decision threshold 0.5 we obtained precision 0.99 and recall 0.92. Method based on connected components patches is less robust: it generate many false positive predictions. But this approach has great potential and for it stable work required much more training examples. Thus convolutional network learned on connected components patches is prominent field for further research.

By using sliding window patches with deep network method we received that mean probability of al-Maqrizi authorship over all al-Khitat pages is equal to 0.86.

V. Conclusion

The join research on al-Maqrizis “Description of Egypt” undertaken by a historian-philologist and three mathematicians from St. Petersburg State University is a unique experiment in working across disciplinary boundaries to achieve a common goal. Its results 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. Given the propensity of Muslim scribes and later writers to attribute manuscripts to important luminaries of the past (such as, e.g., al-Ghazali, d. 505/1111; Ibn al-Arabi, d. 638/1240, and others), the new methods of analyzing and verifying handwritten texts, which have been designed and tested by St. Petersburg mathematicians, are bound to become an important tool for their colleagues in the humanities and social sciences.

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] D. Fecker, A. Asi, W. Pantke, V. Mrgner, 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.

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

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

[6] 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.

[7] 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.

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

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

[10] 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.

[11] 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.