Writer Identification and Retrieval using a Convolutional Neural Network

Stefan Fiel and Robert Sablatnig

Computer Vision Lab
TU Wien
Vienna, Austria
{fiel,sab}@caa.tuwien.ac.at

Abstract. In this paper a novel method for writer identification and retrieval is presented. Writer identification is the process of finding the author of a specific document by comparing it to documents in a database where writers are known, whereas retrieval is the task of finding similar handwritings or all documents of a specific writer. The method presented is using Convolutional Neural Networks (CNN) to generate a feature vector for each writer, which is then compared with the precalculated feature vectors stored in the database. For the generation of this vector the CNN is trained on a database with known writers and after training the classification layer is cut off and the output of the second last fully connected layer is used as feature vector. For the identification a nearest neighbor classification is used. The evaluation is performed on the *ICDAR2013 Competition on Writer Identification*, *ICDAR 2011 Writer Identification Contest*, and the *CVL-Database* datasets. Experiments show, that this novel approach achieves better results to previously presented writer identification approaches.

Keywords: writer identification, writer retrieval, convolutional neural networks

1 Introduction

Writer identification is the task of identifying an author of a handwritten document by comparing the writing with the ones stored in a database. The authors of the documents in the database have to be known in advance for identification. For writer retrieval the documents with the most similar handwriting are searched, generally these are the documents which are written by the same writer. For this task a feature vector is generated, which describes the handwriting of the reference document and the distance to the precalculated features vectors of all documents in the dataset is calculated. For retrieval the documents are sorted according to the distance and for identification the writer of the document with the highest similarity (resp. the smallest distance) is then assigned as author to the document. Writer identification can be used for tasks in forensics like for threat letters, where the writing has to be compared with older ones so

that connections between different letters can be established. Also for historical document analysis writer identification ca be used to trace the routes of medieval scribes along the different monasteries and scriptorias, or to identify the writer of books or pages where the author is not known. Since often a database of known writers is not available for such tasks, the main goal of this approach is to perform writer retrieval. Thus, only a nearest neighbor classification is carried out which allows for searching of documents which have a similar handwriting as a reference document. Especially for the two tasks mentioned the last decision will be made by human experts, but automated methods can be used to reduce the possible handwritings which have to be examined.

The challenges for writer identification and writer retrieval include the use of different pens, which changes a person's writing style, the physical condition of the writer, distractions like multitasking and noise, and also that the writing style changes with age. The changing of the style with increasing age is not covered by any available dataset and cannot be examined, but makes the identification or retrieval harder for real life data. Fig. 1 shows a sample image of the CVL dataset in which the handwriting changes due to distraction of the writer. Fig. 2 shows a part of an image of the CVL dataset, where the writer used two different pens.

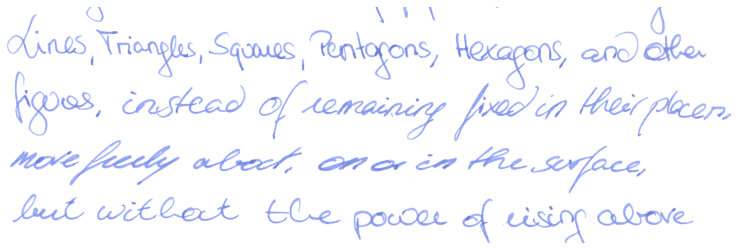


Fig. 1: Part of a sample image of the CVL dataset, which shows the changing of the handwriting when the writer is distracted (Writer Id: 191 Page: 1).

Current state-of-the-art methods either analyze the characters itself by describing their characteristics and properties which are then integrated into a feature vector. Since a binarization and segmentation of the text is necessary, the methods are dependent on these preprocessing steps. To overcome these problems other approaches consider the handwriting as texture and thus use texture analysis methods for the generation of a feature vector. Recent approaches[2][4][5][9] use local features for the task of writer identification which originate in the field of object recognition.

This paper presents an approach that uses Convolutional Neural Networks (CNN) for writer identification and retrieval. CNNs are feed-forward artificial neural networks and are used by currently top ranked methods for object recog-

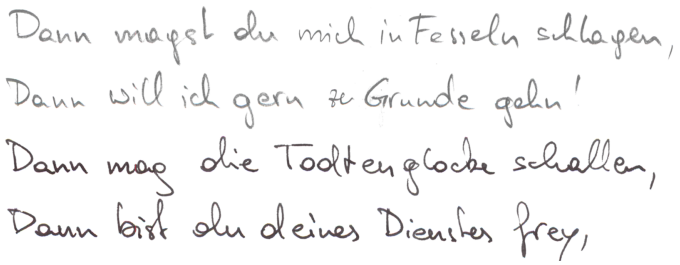


Fig. 2: Part of a sample image of the CVL dataset, on which the writer changed the pen (Writer Id: 369 Page: 6).

nition [18], recognition of digits (MNIST dataset[1]), and speech recognition [19]. They have been brought to the field of text recognition by Wang et al.[20] and are also used for text recognition in natural scenes[16]. To the best of our knowledge, this is the first attempt to bring this method to the field of writer identification and writer retrieval. CNN consists of multiple layers which apply various combinations of convolutions and fully connected neural networks. Since a feature vector is needed for the tasks of identification and retrieval, the last layer of the CNN, which basically does the labeling of the input data, is cut off and the output of the neurons of the second last fully connected layer are used as feature vector. This vector is then used for the distance measurement between two different document images to describe the similarity of the handwriting.

The work is organized as follows: Section 2 gives a brief overview of the current state-of-the-art of writer identification. Section 3 describes the methodology used. The experiments and results are presented in Section 4. Finally, a short conclusion is given in Section 5.

2 Related Work

A writer identification method which is based on features extracted from text lines or characters is proposed by Marti et al. [15]. Features like slant, width, and three heights of the writing zone (descender, x-height, and ascender height) are used for the identification of the writer. Using a neural network classifier, a recognition rate of 90% on 20 writers of the IAM Database is achieved. New features which are calculated on character level are introduced by Bulacu et al. [1]. They calculate the contour-hinge, which describes various angles of the written character. Furthermore they use a writer-specific grapheme emission and the run-length. With a nearest neighbor classifier they achieve a recognition rate of 89% on the Firemaker dataset, which contains of 250 writers. Another approach is the "Grid Microstructure Features" which are introduced by Li and Ding [11]. For each border pixel of the edge of a writing the neighborhood is described by means of three simple rules. These rules describe the characteristics of the edge

[1] http://yann.lecun.com/exdb/mnist/ - accessed March 2015

on the connected component within a small window. The probability density distribution of different pixel pairs which fullfill these rules is regarded as feature representing the writing style. With this method they were able to win the *ICDAR 2011 Writer Identification Contest*[13] with an identification rate of 99.5% on the complete pages and 90.9% on cropped pages, each of the datasets are written by 26 writers. Jain and Doermann [7] propose an offline writer identification method by using K-adjacent segment features in a bag-of-features framework. It represents the relationship between sets of neighboring edges in an image which are then used to form a codebook and classify new handwritings using the nearest neighbors. A recognition rate of 93.3% is achieved on 300 writers of the IAM Dataset. The same authors propose the usage of an alphabet of contour gradient descriptors for writer identification [8]. By analyzing the contour gradients of the characters in a sliding window, they form a pseudo-alphabet for each writing sample and calculate the distance between these alphabets as similarity measurement. With this method they were able to win the *ICDAR 2013 Competition on Writer Identification*[12] with an identification rate of 95.1%. The dataset was written by 250 writers.

Hiremath et al. [6] assume the writing as texture image and are proposing to use the wavelet transform to compute co-occurrence matrices for 8 directions. When dealing with 30 writers at a time the identification rate is 88%. Our previous approaches [4] [5] use SIFT features for classification. With a bag-of-words approach respectively using the Fisher information of Gaussian Mixture Models an identification rate of 90.8% and 99.5% is achieved on the CVL Dataset with 310 writers. On the same dataset Christlein et al. [2] achieve an accuracy of up to 99.2% by using RootSIFT and GMM supervectors for their identification system. Jain and Doermann [9] propose a combination of local features. They use a linear combination of their k-adjacent segments, their alphabet of contour gradient descriptors, and SURF for identification. With a nearest neighbor approach they achieve a recognition rate of 99.4% on the CVL Dataset.

The evaluation of all methods has been carried out on various databases with different properties and thus the results cannot be compared with another.

3 Methodology

Our approach uses CNN for the task of writer identification and writer retrieval. Since a feature vector is needed for every document image to allow for a comparison with precalculated features in a dataset to identify a specific writer or to search for the most similar handwriting style, the output of the second last fully connected layer is used. CNNs require as input an image with fixed size, thus preprocessing of the document images is necessary. The preprocessing includes binarization, text line segmentation, and sliding windows. The next step is the generation of the feature vector using the CNN. These vectors are then used for the identification of a writer or the retrieval of similar writers using a nearest neighbor approach.

3.1 Preprocessing

The first step is, if the input images are grayscale like in the CVL-Dataset [10], a binarization. For this work the method of Otsu [17] is used since the dataset contains only scanned pages without any noise and thus a global threshold will give a nearly optimal binarization. Second, the words respectively the lines have to be segmented. The CVL-Database [10] and the IAM-Dataset [14] already provide a segmentation of the words, thus these images are used for evaluation and training. For the *ICDAR 2011 Writer Identification Contest*[13] and *ICDAR 2013 Competition on Writer Identification* [12] datasets the lines are segmented using the method of Diem et al.[3], which uses Local Projection Profiles for grouping the characters to words. The text lines are then detected by globally minimizing the distances of all words. Since the CNNs require an image of fixed size as input, the word images respectively the line images are split up using a sliding window approach with a step size of 20 pixels. But first these images are size normalized to ensure that the height of the writing does not have an influence on the feature vector generation. Thus, the x-height of the words or lines are calculated by fitting a line through the upper and lower points of the text line and the image is resized that the x-height of the words cover half of the result image, to ensure that ascenders and descenders have sufficient space to be also present in the image. Additionally, since some lines in the ICDAR datasets are slightly skewed, the lines are also deskewed with the mean angle of the upper and lower profile of the x-height. The upper image in Fig. 3 shows the original line from the ICDAR 2011 dataset with the profiles of the x-height as new line. The lower image shows the deskewed line, which was also size normalized. Note that not the slant of the font is corrected since it is a discriminative feature of the writer, only the orientation of the text line is processed which cannot be used as feature for identification since it is highly depended on the paper the text is written on, e.g. if it is a lined or blank paper.

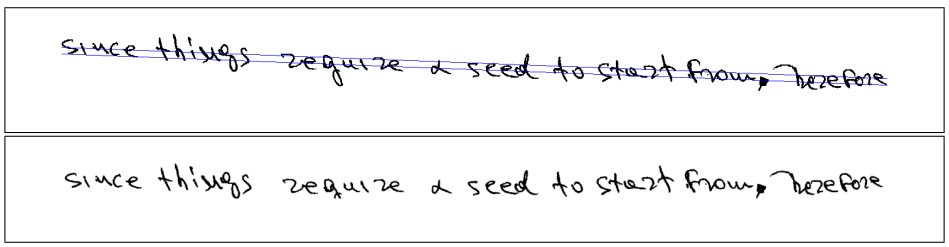


Fig. 3: Sample line extracted from the ICDAR 2011 dataset. The upper image shows the original line with the upper and lower profile. The lower images shows the size normalized and deskewed line.

3.2 Generation of the Feature Vector

For the generation of the feature vector a CNN is used. For this work a well-known model for CNN is used, namely "caffenet" which is part of the "Caffe - Deep learning framework"[2]. The design of the network is presented in Fig. 4 . It consists of five convolution layers which are using kernel sizes of 11-3 and three fully connected layers. Like the original network it is trained using a softmax loss function. The last layer of the CNN is used for labeling the input data and consists of 1000 neurons, which is more than the actual number of writers in the IAM dataset but leads to better results. The reason for this behavior will be examined in more detail as future work. Since a classification is not intended, this layer is cut off and the output values of the second last fully connected layer is used as feature vector for further processing. The classification layer could have been also used as feature vector for the writer identification but lead to worse results since the outputs of this layer focuses rather on one class whereas when using the last fully connected layer all neurons are activated and thus giving a more discriminative feature vector.

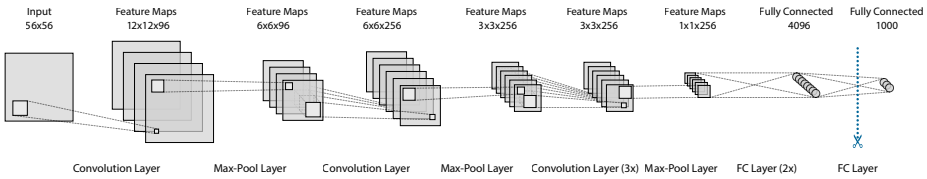


Fig. 4: Design of the CNN, the "caffenet" of the "Caffe - Deep learning framwork"

The CNN has to be trained beforehand. To ensure independence between the training images and the ones used for evaluation, the CNN has been trained on the word images of the IAM dataset. The IAM Dataset consists of 1539 document images written by 657 different writers. The document images are not equally distributed among the writers, most of the writers only contributed one document whereas one writer has written 60 pages. Since CNNs have to be trained on a large amount of data to achieve a good performance (e.g. for ILSVCRC 2014 the training set contained 1.2 Mio images), the trainings set has been enlarged artificially by rotating each image of a sliding window from -25 to $+25$ degrees using a step size of 5 degrees. These values are found empirically and are a good trade-off between the performance of the method and training time. The rotation of the images may also have a positive effect for handwritings with a certain slant, these images are no longer assigned to the same writer in the training dataset using a similar slant. This property has to be confirmed in future work. With the rotation of the image the training set consists of more than 2.3 Mio image patches, which are not equally distributed among the 657

[2] http://caffe.berkeleyvision.org - accessed March 2015

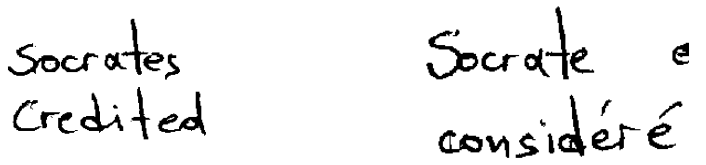
writers of the dataset due to their properties. Each writer has at most 7700 patches (700 for each direction) in the trainings set.

To generate a feature vector for a complete page, all image patches of this page are fed into the CNN without any rotation since experiments have shown that the performance is not improved. As mentioned above, the last layer is cut off for the generation of the feature vector. Thus, we receive the normalized output of the last 4096 neurons of the second last fully connected layer. The mean values of the vectors of all image patches of one page is then taken as feature vector for the identification respectively the retrieval. These feature vectors are compared with each other using the χ^2-distance, which has been found out empirically to give the best results.

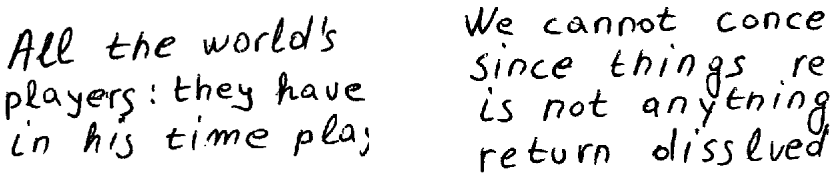
4 Experiments and Results

This section will give an overview of the performance of the method presented on various databases. For the evaluation the datasets of the ICDAR 2011 and ICDAR 2013 Writer identification contests, as well as the CVL database have been used. The ICDAR 2011 dataset consists of 26 writers, where each has contributed 8 different documents (two in English, French, German, and Greek). The second dataset of this contest contains the same pages, but only the first two text lines from each document are taken. Fig. 5a show two small parts of images of this dataset. The ICDAR 2013 dataset contains the document images from 50 writers, one in English and one in Greek. From each image two pieces were cropped, each containing four text line, thus resulting in four parts of text per writer. Two parts of sample images are shown in Fig. 5b. The CVL dataset is the largest dataset in this evaluation. It consists of 1545 pages written by 309 writers. Each writer has written 5 different texts (four in English, one in German). Sample images have already been presented in Section 1. As mentioned in Section 3.2 the CNN has been trained on the IAM database to ensure independence of the trainings set and the evaluation sets. One CNN has been trained for all experiments although different designs of the CNN slightly improved the performances for some experiments. The results of the CNN which has the best overall performance are presented.

The evaluation has been carried out in the same way as in the ICDAR 2011 and ICDAR 2013 contest. For each document a ranking of the other documents according to the similarity is generated. There the top N documents are examined whether they are from the same writer or not. Two criteria have been defined: for the soft criteria, if one of the documents in the top N is from the same writer like the reference document it is considered as a correct hit. For the hard criterion all N documents have to be from the same writer to be considered as correct hit. The value of N is varying for all the datasets, since the number of documents from one writer is also varying. For the ICDAR 2011 contest the values of N are: 1, 2, 5, and 7, whereas the values for the hard criterion are 2, 5, and 7. For the ICDAR 2013 dataset the values for the soft criterion are 1, 2, 5, and 10 and for the hard criterion 2 and 3. The CVL dataset uses the same num-



(a) ICDAR 11, WriterId: 2, Text 2 and Text 3



(b) ICDAR 13, Writer: 29, Text 1 and Text 2

Fig. 5: Parts of sample images of the ICDAR 2011 and the ICDAR 2013 dataset.

bers for the soft criterion and for the hard criterion 2, 3, and 4 are used, since the dataset contains 5 documents of each writer. Furthermore the CVL dataset has a retrieval criterion, which is defined as the percentage of the documents of the corresponding writer in the first N documents. For this criterion the values of N are the same as for the hard criterion.

The first evaluation is carried out on the ICDAR 2011 datasets. The results for the soft criterion of both datasets are presented in Table 1 in comparison with the three best ranked methods of the contest. It can be seen that for the soft criterion all methods have a good performance, our method has one misclassified page in the "Top 1" task. On the cropped dataset the performance of all methods are dropping, since there is less written text in the image. Still our method has the best performance for the "Top 1", which is the exact identification of the writer.

The evaluation of the hard criterion of the ICDAR 2011 datasets are shown in Table 2. For the hard criterion similar results can be seen like for the soft criterion. The proposed method outperforms the other methods for all but one task. For the "Top 2" task on the cropped dataset, the improvement is 4.8%. Only the results of the "Top 7" task on the cropped dataset, which is finding all other 7 pages of the same writer, the proposed method provides slightly worse results than the other methods.

The next experiments have been carried out on the ICDAR 2013 dataset. The results of both criteria compared to the top ranked methods of the contests are presented in Table 3. It can be seen that the proposed method performs worse than the best three participants of the competition. This has two reasons: the line segmentation used has problems on this dataset and often the ascenders and descenders of characters are cut off and thus these characteristic parts for different writers are missing in the classification. Second, like the "CS-UMD"

Table 1: The soft criterion evaluation results on the ICDAR 2011 dataset (in %)

complete dataset				
	Top 1	Top 2	Top 5	Top 7
Tsinghua	98.6	100.0	100.0	100.0
MCS-NUST	99.0	99.5	99.5	99.5
Tebessa C	98.6	100.0	100.0	100.0
proposed method	**99.5**	100.0	100.0	100.0
cropped dataset				
	Top 1	Top 2	Top 5	Top 7
Tsinghua	90.9	93.8	**98.6**	**99.5**
MCS-NUST	82.2	91.8	96.6	97.6
Tebessa C	87.5	92.8	97.6	**99.5**
proposed method	**94.7**	**97.6**	98.1	**99.5**

Table 2: The hard criterion evaluation results on the ICDAR 2011 dataset (in %)

complete dataset			
	Top 2	Top 5	Top 7
Tsinghua	95.2	84.1	41.4
MCS-NUST	93.3	78.9	39.9
Tebessa C	97.1	81.3	50.0
proposed method	**98.6**	**87.0**	**52.4**
cropped dataset			
	Top 2	Top 5	Top 7
Tsinghua	79.8	48.6	12.5
MCS-NUST	71.6	35.6	11.1
Tebessa C	76.0	34.1	**14.4**
proposed method	**84.6**	**53.8**	10.1

methods, the proposed method has difficulties in finding the corresponding Greek text for an English input text and vice versa since the training data does not contain any Greek text. This can be seen in the hard criterion, where at least one document image written in the other language has to be found.

Table 3: Evaluation of the soft and hard criteria on the ICDAR 2013 dataset (in %)

	soft criterion				hard criterion	
	Top 1	Top 2	Top 5	Top 10	Top 2	Top 3
CS-UMD-a	**95.1**	**97.7**	98.6	99.1	19.6	7.1
CS-UMD-b	95.0	97.2	**98.6**	**99.2**	20.2	8.4
HIT-ICG	94.8	96.7	98.0	98.3	**63.2**	**36.5**
proposed method	88.5	92.2	96.0	98.3	40.5	15.8

The last evaluation has been carried out on the CVL dataset. The results of the soft and hard criterion are shown in Table 4, whereas results of the retrieval criterion are shown in Table 5. The proposed method has the best performance on each task except for the "Top 3" of the hard criterion compared to the top ranked methods in [10]. Remarkable is the improvement of performance for "Top 4" hard criterion, which is 6.9%. For this task all other 4 pages of one writer has to be retrieved. These results can also be seen in the retrieval criterion, where the proposed method achieves higher results as the other methods.

Table 4: Evaluation results of the soft and hard criteria on the CVL-Database (in %)

	soft criterion				hard criterion		
	Top 1	Top 2	Top 5	Top 10	Top 2	Top 3	Top 4
Tsinghua	97.7	98.3	99.0	99.1	95.3	**94.5**	73.0
Tebessa C	97.6	97.9	98.3	98.5	94.3	88.2	73.0
proposed method	**98.9**	**99.0**	**99.3**	**99.5**	**97.6**	93.3	**79.9**

Table 5: The retrieval criterion evaluation results on the CVL-Database (in %)

	Top 2	Top 3	Top 4
Tsinghua	96.8	94.5	90.2
Tebessa C	96.1	94.2	90.0
proposed method	**98.3**	**96.9**	**93.3**

5 Conclusion

A novel method for writer identification and writer retrieval has been presented. The method uses CNN for generating a feature vector which are then compared using the χ^2-distance. As preprocessing steps the images need to be binarized, normalized, and the skew of the text line needs to be corrected. The method proposed has been evaluated on three different datasets, namely the datasets of the ICDAR 2011 and 2013 writer identification contests and the CVL dataset. Experiments show that the proposed method achieves slightly better results on two of three datasets, but worse results on the remaining dataset which originates mainly from the preprocessing steps.

Future work includes the design of a new CNN customized to the input data and which is capable of achieving better performance on various datasets. Furthermore the preprocessing step will be improved by using a better normalization of the image patches and a post processing step with a voting strategy on the

complete page and the rejection of not significant image patches will be introduced. Also some image patches may be skipped already in the preprocessing step if they show no relevant information for a successful writer identification and writer retrieval.

References

1. Bulacu, M., Schomaker, L., Vuurpijl, L.: Writer identification using edge-based directional features. In: Proceedings. Seventh International Conference on Document Analysis and Recognition, 2003. pp. 937 – 941 (Aug 2003)
2. Christlein, V., Bernecker, D., Hönig, F., Angelopoulou, E.: Writer Identification and Verification Using GMM Supervectors. In: Proceedings of the 2014 IEEE Winter Conference on Applications of Computer Vision (2014)
3. Diem, M., Kleber, F., Sablatnig, R.: Text Line Detection for Heterogeneous Documents. In: 2013 12th International Conference on Document Analysis and Recognition (ICDAR). pp. 743–747 (2013)
4. Fiel, S., Sablatnig, R.: Writer Retrieval and Writer Identification Using Local Features. In: 2012 10th IAPR International Workshop on Document Analysis Systems (DAS). pp. 145 –149. IEEE (march 2012)
5. Fiel, S., Sablatnig, R.: Writer Identification and Writer Retrieval Using the Fisher Vector on Visual Vocabularies. In: 2013 12th International Conference on Document Analysis and Recognition (ICDAR). pp. 545–549 (2013)
6. Hiremath, P., Shivashankar, S., Pujari, J., Kartik, R.: Writer identification in a handwritten document image using texture features. In: International Conference on Signal and Image Processing (ICSIP). pp. 139 –142 (dec 2010)
7. Jain, R., Doermann, D.: Offline Writer Identification Using K-Adjacent Segments. In: 2011 International Conference on Document Analysis and Recognition (ICDAR). pp. 769 –773 (sept 2011)
8. Jain, R., Doermann, D.: Writer Identification Using an Alphabet of Contour Gradient Descriptors. In: 2013 12th International Conference on Document Analysis and Recognition (ICDAR). pp. 550–554 (Aug 2013)
9. Jain, R., Doermann, D.: Combining Local Features for Offline Writer Identification. In: 2014 14th International Conference on Frontiers in Handwriting Recognition (ICFHR). pp. 583–588 (Sept 2014)
10. Kleber, F., Fiel, S., Diem, M., Sablatnig, R.: CVL-DataBase: An Off-Line Database for Writer Retrieval, Writer Identification and Word Spotting. In: 2013 12th International Conference on Document Analysis and Recognition (ICDAR). pp. 560–564 (2013)
11. Li, X., Ding, X.: Writer Identification of Chinese Handwriting Using Grid Microstructure Feature. In: Advances in Biometrics, Lecture Notes in Computer Science, vol. 5558, pp. 1230–1239. Springer Berlin / Heidelberg (2009)
12. Louloudis, G., Gatos, B., Stamatopoulos, N., Papandreou, A.: ICDAR 2013 Competition on Writer Identification. In: 2013 12th International Conference on Document Analysis and Recognition (ICDAR). pp. 1397–1401 (Aug 2013)
13. Louloudis, G., Stamatopoulos, N., Gatos, B.: ICDAR 2011 Writer Identification Contest. 2011 11th International Conference on Document Analysis and Recognition (ICDAR) pp. 1475–1479 (2011)
14. Marti, U.V., Bunke, H.: The IAM-database: an English sentence database for offline handwriting recognition. International Journal on Document Analysis and Recognition 5(1), 39–46 (2002)

15. Marti, U.V., Messerli, R., Bunke, H.: Writer identification using text line based features. In: Proceedings. Sixth International Conference on Document Analysis and Recognition. pp. 101 –105 (2001)
16. Opitz, M., Diem, M., Fiel, S., Kleber, F., Sablatnig, R.: End-to-End Text Recognition with Local Ternary Patterns, MSER and Deep Convolutional Nets. In: Proceedings of the 11th International Workshop on Document Analysis Systems. pp. 186–190 (2014)
17. Otsu, N.: A Threshold Selection Method from Gray-Level Histograms. IEEE Transactions on Systems, Man and Cybernetics 9(1), 62–66 (Jan 1979)
18. Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M.S., Berg, A.C., Fei-Fei, L.: ImageNet Large Scale Visual Recognition Challenge. CoRR abs/1409.0575 (2014), http://arxiv.org/abs/1409.0575
19. Sainath, T., Mohamed, A.R., Kingsbury, B., Ramabhadran, B.: Deep convolutional neural networks for LVCSR. In: 2013 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). pp. 8614–8618 (May 2013)
20. Wang, T., Wu, D., Coates, A., Ng, A.: End-to-end text recognition with convolutional neural networks. In: 2012 21st International Conference on Pattern Recognition (ICPR). pp. 3304–3308 (Nov 2012)