Writer identification for historical handwritten documents using a single feature extraction method

Michel Chammas, Abdallah Makhoul, Jacques Demerjian

Writer identification for historical handwritten documents using a single feature extraction method

Michel Chammas*†, Abdallah Makhoul† and Jacques Demerjian‡
*Digital Humanities Center, University of Balamand, El-Koura, Lebanon
michel.chammas@balamand.edu.lb
†Femto-ST Institute, UMR CNRS 6174, Université de Bourgogne Franche-Comté, Montbéliard, France
Email: michel.chammas@univ-fcomte.fr, abdallah.makhoul@univ-fcomte.fr
‡LaRRIS, Faculty of Sciences, Lebanese University, Fanar, Lebanon
jacques.demerjian@ul.edu.lb

***Abstract*—With the growth of artificial intelligence techniques the problem of writer identification from historical documents has gained increased interest. It consists on knowing the identity of writers of these documents. This paper introduces our baseline system for writer identification, tested on a large dataset of latin historical manuscripts used in the ICDAR 2019 competition. The proposed system yielded the best results using Scale Invariant Feature Transform (SIFT) as a single feature extraction method, without any preprocessing stage. The system was compared against four teams who participated in the competition with different feature extraction methods: SRS-LBP, SIFT, Pathlet, Hinge, Co-Hinge, QuadHinge, Quill, TCC and oBIFs. An unsupervised learning system was implemented, where a deep Convolutional Neural Network (CNN) was trained using patches extracted from SIFT descriptors. Then the results were encoded using a multi - Vector of Locally Aggregated Descriptors (VLAD) and applied an Exemplar Support Vector Machine (E-SVM) at the end to compare the results. Our system achieved best performance using a single feature extraction method with 91.2% mean Average Precision (mAP) and 97.0% accuracy.**

***Index Terms*—Writer identification, historical documents, artificial intelligence, sift descriptors.**

I. Introduction

Knowing the identity of the writer in historical manuscripts is one of the most important challenges for humanitarians [1]. They depend on many specific characteristics to determine the typography and the era of the text. One of their main concerns is to identify or predict the scribe of the text, where most of the time stay uncertain. In many manuscripts, the writer's name is mentioned at the last page in a small paragraph called colophon [2], through which they could recognize the identity of the writer. They will be able to compare the writing style of the text to other manuscripts, which belong to the same writer. This work requires a considerable amount of time and effort. Whilst the development of automated learning techniques facilitated the prediction of writer identity, but this topic remains a challenge with the difficulty of reading some texts or detecting their origins [3]. We developed a system based on machine learning algorithms to identify writers of historical handwritten manuscripts. Our main concern was to make sure that our system generalizes well on test data. Therefore, we decided to validate our system and assess its performance on a public dataset.

Many international competitions were done on this topic: ICFHR2016 Competition on the Classification of Medieval Handwritings in Latin Script (CLAMM) [4], ICDAR2017 CLAMM [5] and the latest was ICDAR2019 Competition on Image Retrieval for Historical Handwritten Documents (ICDAR-2019-HDRC-IR) [6].

We evaluated our system on the leading and well-known competition ICDAR 2019 dataset [6]. A total of four teams participated in the competition with multiple submissions of different methods. The system of the first team, from the Universitat Autònoma de Barcelona (UAB), was used as a baseline system based on the Sparse Radial Sampling of Local Binary Patterns (SRS-LBP). The second team from the South China University of Technology (SCUT) participated with two different methods submissions as follows: SIFT with Fisher vector encoding and Pathlet with VLAD encoding. The third team from the University of Groningen submitted five methods: Hinge, Co-Hinge, QuadHinge, Quill and TCC. While the fourth team from the University of Tebessa submitted only one method using oriented Basic Image Features (oBIFs).

In the proposed system, we used SIFT as feature extraction method along with CNN to train the system using patches extracted from SIFT descriptors. Then we encoded the results using multi-VLAD (with 5 layers) and applied an exemplar SVM at the end to compare the results. We used l2-norm to normalize the data at each stage. We yielded the best results with 91% mAP and 96.9% accuracy. The result of each method will be presented in the evaluation section, where we compare all the results of all mentioned methods with our system.

The rest of this paper is organized as follows: Section II presents briefly the related works in this field, Section III shows the dataset used and how the images are divided, Section IV describes the proposed system and its functionality, Section V evaluates the system results and

978-1-7281-8470-8/20/$31.00 ©2020 IEEE
DOI 10.1109/ICMLA51294.2020.00010

compares them with the competition results, and Section VI concludes the paper.

II. Related works

This section briefly reviews related works to writer identification. Previous research worked initially on the techniques of extracting writer-specific features from handwritten patterns. Several methods were used for feature extraction like SIFT, RootSIFT, Contour, Hinge, Path Signature and many more handcrafted techniques [7] [8]. Afterward, the extracted features descriptors were either classified using traditional classifiers such as distance-based classifier, Support Vector Machines (SVM), Hidden Markov Model (HMM) and Fuzzy based classifier, or trained using a deep neural network [9]. Neural Networks were also used for automatic feature extraction [10].

Most of the previous techniques were based on supervised learning through a single or a combination of multi handcrafted feature extractors. He. et al used handcrafted features by computing the junctions of the handwriting along with SVM. But they used clustering to create an unsupervised approach, which proved better features extraction [11]. However, deep neural network, specially CNN, showed superior performance as unsupervised learning [12].

Fiel and Sablatnig used CNN to automatically extract the features vector. In their approach, they eliminated the last layer of the network and computed the distance using ChiSquare to identify the writers [10].

Christlein et al. used CNN along with SIFT for feature extraction and encoded the result vector with Gaussian Mixture Models (GMM) [13]. According to Christlein, CNN achieved better performance when using the SIFT descriptors as a local feature extractor [12].

Lai et al. introduced a different approach based on path signature, where they extract the pathlets from the polygonized handwriting contours. This technique requires first to binarize the image and extract the contour of the text. Then the extracted features are clustered using k-means to create a codebook [8].

Nguyen et al. used CNN end-to-end deep-learning method to extract local features, then they aggregated the results in a global feature vector. After that, they used a softmax classifier to predict the writers without requirement of prior identification of features [9].

Based on those studies and experiments, we concluded that using SIFT in conjunction with CNN as feature extraction method would acquire the best results. We improved the approach of Christlein [12] by training a CNN to extract local features from image patches. The SIFT was used to extract keypoint decriptors and image patches, where the descriptors were clustered and used to label the image patches. The local features extracted by the CNN network were used to create a VLAD codebook, which was encoded and normalized to perform as a global descriptor. In the next section we will describe the dataset used in our proposal.

III. Dataset

The ICDAR dataset consists of a large number of Latin historical handwritten documents. The training data contains 300 writers contributing 1 page, 100 writers contributing 3 pages, and 120 writers contributing 5 pages resulting in 1200 images of 520 writers. And the testing data contains 20,000 images, about 7 500 pages stem from isolated documents (partially anonymous writers, contributing one page each), and about 12 500 pages are from writers that contributed three or five pages. Fig. 1 shows two samples of Manuscripts used in the dataset.

The metrics considered during the competition are the mean Average Precision (mAP) and Accuracy [6]. TABLE I shows more details (origins, number of writers and number of images per writer) about the dataset used in ICDAR 2019 [6]. The dataset is publicly available at the following URL: https://doi.org/10.5281/zenodo.3262372.

IV. System description

In order to perform writer identification for historical handwritten manuscripts, two steps are required: feature extraction and classification.

A. Features extraction

The unsupervised Scale Invariant Feature Transform (SIFT) [12] method is used to localize potential image samples for feature extraction. SIFT method has the advantage of being invariant to basic transformations such as scaling, rotation and translation. Centered at each keypoint, a 32x32 patch is extracted from the document image. Each patch is mapped to a 128 dimensional SIFT descriptor vector. The importance of a SIFT descriptor is that it is invariant to transformation (scaling, rotation and translation). Fig. 2 shows a sample of the extracted SIFT keypoints and their descriptors.

In order to improve the recognition performance and obtain a compact representation, the Principal Component Analysis (PCA) method (with whitening) [14] is applied on the SIFT descriptors in order to reduce their dimensionality. We reduce the descriptors dimension from 128 to 32. This reduction helps to decrease the processing cost of the descriptor clustering operation.

B. Clustering

We used the mini batch k-means clustering technique [13] (with initial size of 15000) to group the descriptors into 5000 clusters. The clusters are used as labels for image patches. We randomly extracted 990k patches with their respective labels (cluster centroids) from the training data (275 patch per document). These patches were used to train a deep Convolutional Neural Network (CNN) to be used as a feature extractor [15]. The CNN we used is a Residual Neural Network with 20 layers. To train the network, we used 900k patches for training and 90k for testing.

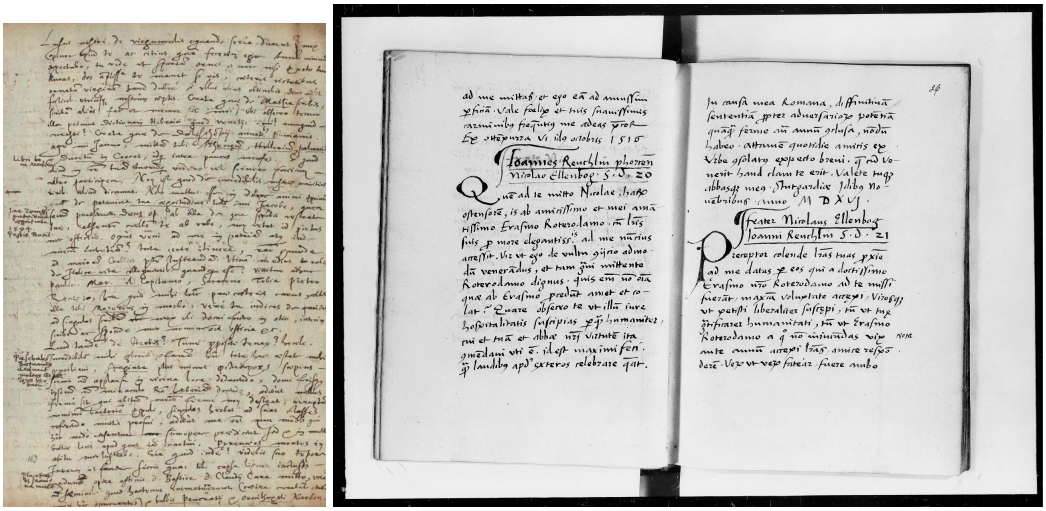

Fig. 1: The image on the left represents a sample page of the training data and the image on the right represents a sample page of the testing data

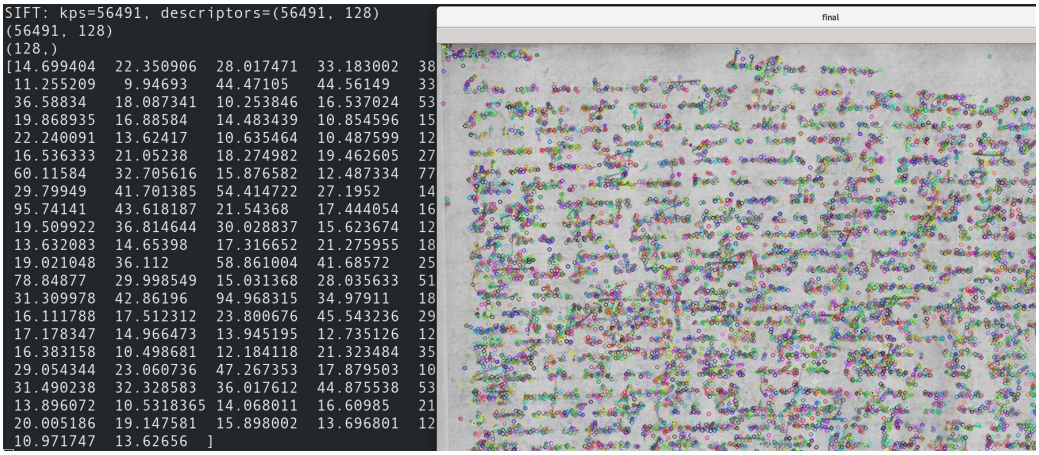

Fig. 2: SIFT keypoints (on the right) and a descriptor vector of a keypoint (on the left), for example on this image sample 56491 keypoints are extracted and a vector of 128 descriptor at each keypoint

Local descriptors (of dimension 64) extracted by CNN from the document of the training data were used to create a Universal Background Model (UBM). The UBM is obtained via k-means clustering with 100 centroids. The number of clusters, patches and centroids were chosen based on our readings and previous experiments on different datasets.

C. CNN training

We trained a CNN to map image patches into their corresponding labels. The CNN is then used as a feature extractor [15]. We split the data for training and validation, where 90k are used for validation. We implemented a deep residual network with 20 layers (ResNet20), where local descriptors were extracted with CNN for each document [10]. Fig. 3 shows the accuracy rating of the ResNet20 training. After extracting the CNN features, we created a UBM from all the descriptors in the training data and clustered them using k-means to 100 clusters. Then we applied l2-norm for data normalisation [12].

$$l2(v) = ||v||2 \tag{1}$$

$$||v||2 = \sqrt{a1^2 + a2^2 + a3^2} \tag{2}$$

D. Encoding

A VLAD codebook, acting as a Universal Background Model [1], is formed via k-means clustering of all the local descriptors in the training data into a set of 100 clusters. The VLAD codebook is then used to encode each document [16]. For instance, local descriptors of each document are

TABLE I: Image providers and number of writers used in the test dataset [6]

Provider	City	nb writers	nb images/writer	images total
Manuscripts		2027		10135
Bodleian Libr.	Oxford	9	5	
BVMM		586	5	
	Boulogne	28		
	Chantilly	30		
	Nantes	13		
	Rennes	16		
	Saint-Omer	363		
	Toulouse	12		
	10 writers p. repository	124		
Cambridge Dig. Libr.	2	5		
e-codices	Geneva	2	5	
Gallica		1352	5	
	Amiens	14		
	Paris	1232		
	Reims	41		
	Valenciennes	52		
	8 writers p. repository	13		
Harvard		19	5	
Stanford	Baltimore (Walters)	57	5	
Letters A		831		2655
Univ. Library	Erlangen			
– Erlangen-Nurnberg		290	1	
		170	3	
		371	5	
Letters B		2052		2052
Univ. Library Basel	Basel	2052	1	2052
Charters		5158		5158
Monasterium		5158	1	5158
Total		**10068**		**20000**

aggregated into a 6400 dimensional global descriptor vector. Then we applied l2 norm to normalize the VLAD vectors [17].

To further improve performance, we used the multi-VLAD approach in which we created 5 different codebooks and concatenated their global encodings into a 32000 dimensional global descriptor. Finally, we reduce the global descriptor dimensionality to 3200 via regularized PCA [16].

$$\Sigma = \frac{1}{n-1}\left((\mathbf{X}-\bar{\mathbf{x}})^T\,(\mathbf{X}-\bar{\mathbf{x}})\right) \tag{3}$$

$$\bar{\mathbf{x}} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{4}$$

E. Similarity

At the end, we compared the documents by computing the cosine distance between their VLAD vectors. Than we implemented Exemplar-SVM (ESVM) using a linear kernel [18] [12], as a result we reached an improved precision and accuracy.

V. Evaluation

The system of the first team from UAB was used as a baseline system based on the method of Nicolaou et al. [19], the Sparse Radial Sampling of Local Binary Patterns (SRS-LBP) and Hellinger Kernel, they achieved 86.6% mAP and 93.1% accuracy for the retrieval using Manhattan to compute the distance [6]. The second team from the South China University of Technology (SCUT) participated with three submissions, one with two combined features (SIFT and Pathlet) and two with a single feature extraction method as follows: SIFT with Fisher vector encoding and Pathlet with VLAD encoding. On the first stage before feature extraction, they performed two pre-processing steps. The first step is a text binarization using deep Unet, while the second step is a page-level rotation Correction to align the text. In their

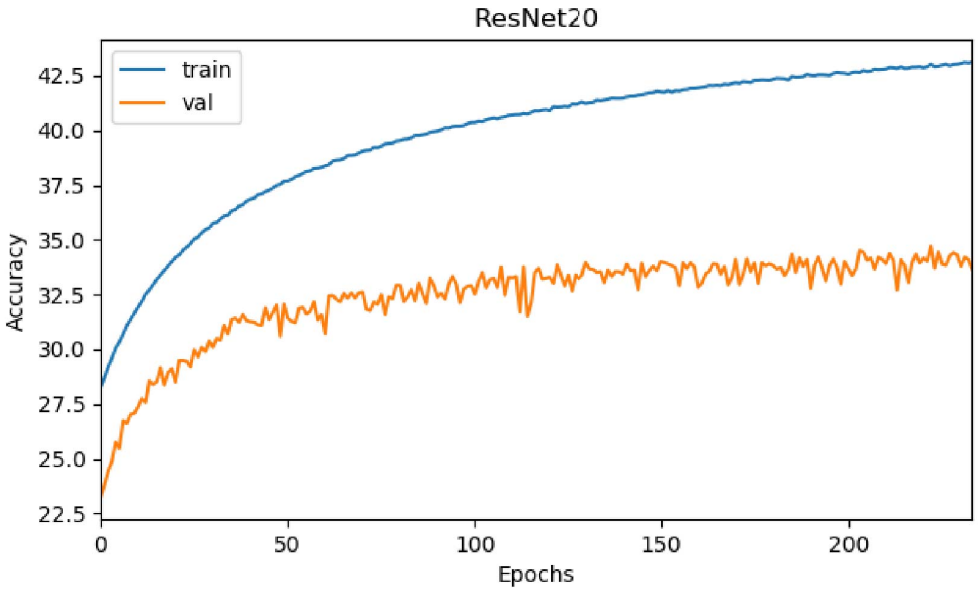

Fig. 3: ResNet20 training and validation accuracy

first submission, they extracted SIFT features and used the Fisher Vector method to encode them. They achieved their best result with 90.6% mAP and 96.6% accuracy. In their second submission, they extracted Pathlet features and encoded them with VLAD. They got 89.8% mAP and 96% accuracy [6]. The third team from the University of Groningen also applied binarization as a pre-processing step and used X2 distance to compute the scores. They submitted five methods: Hinge (75.6% mAP and 88.4% accuracy), Co-Hinge (84.5% mAP and 92.9% accuracy), QuadHinge (80.2% mAP and 91.3% accuracy), Quill (76% mAP and 88.3% accuracy) and TCC (79% mAP and 89.7% accuracy) [6]. The fourth team from the University of Tebessa submitted only one method using oriented Basic Image Features (oBIFs) without any preprocessing and used Euclidian distance to calculate the result: 84.6% mAP and 92.7% accuracy [6]. TABLE II shows the results of all submissions (teams, methods and scores).

In contrast to the other systems, our system works directly on raw images where no preprocessing step is involved and uses a neural network-based approach. We used a single method: for feature extraction (SIFT) and trained a CNN using the SIFT descriptors. We encoded the results using VLAD and computed the similarity scores using cosine distance (89.7% mAP and 95.5% accuracy). Then we tried to compute the distance using ESVM (linear kernel) and we succeeded to improve the results by around 1.5%. We got the best result compared to similar submission of the SCUTT team using SIFT with 91.2% mAP and 97.0% accuracy. The importance of our system is that we achieved a better result to the same method without any preprocessing step. We applied SIFT, which was implemented by Christlein et al. in 2017 [13], to extract the features directly from the images provided in the dataset. While the SCUTT team first used a deep Unet for text binarization, which can achieve a high precision recall and improve the accuracy performance by 1-2% [20]. Second, they performed a page-level rotation correction step based on the line projection method in order to make the text more horizontal. They used the extracted binary and gray-scale images to perform the feature extraction in their both methods. The SCUTT team won the ICDAR 2019 competition as they achieved the highest mAP with two combined methods SIFT and Pathlet [6], but we were able to yield the better mAP and accuracy with our system using a single method.

VI. Conclusion

In this paper, we presented a performant system for writer identification using only one feature extraction method and without any pre-processing techniques. Our system presents best results compared to other participants. It is notable that SIFT is the best method for feature extraction as it surpassed all the other methods. Also, we showed that ESVM improved our results when used as a replacement of the cosine distance to compute the distance between the VLAD vectors. For future work, we will take into consideration the use of two feature extraction methods. Also, since all the competitions were challenged with Latin Handwritten Manuscripts and we did not find any assessed Historical Arabic Handwritten dataset, we will test our system on the historical Arabic manuscripts

TABLE II: The results of ICDAR 2019 competition with single methods compared to our result [6]

	Method	Accuracy [%]	mAP [%]
Baseline	SRS-LBP (a) Classification	92.2	84.0
	SRS-LBP (b) Retrieval	93.1	86.8
SCUTT	SIFT	96.6	90.6
	Pathlet	96.0	89.8
Groningen	Hinge	88.4	75.6
	Co-Hinge	92.9	84.5
	QuadHinge	91.3	80.2
	Quill	88.3	76.0
	TCC	89.7	79.0
Tebessa	oBIFs	92.7	84.6
Our System	**SIFT**	**97.0**	**91.2**

that are available at the Digital Humanities Center at the University of Balamand. The collection contains more than 500 Arabic manuscripts from the 8th to the 19th century. It includes more than 11000 images for more than 500 scribes, and gathered from more than 10 locations between Lebanon and Syria. At the end, we look forward for the next competition.

Acknowledgement

This work has been supported by the EIPHI Graduate School (contract ANR-17-EURE-0002).

References

[1] A. Rehman, S. Naz, and M. I. Razzak, "Writer identification using machine learning approaches: a comprehensive review," *Multimedia Tools and Applications*, vol. 78, no. 8, pp. 10 889–10 931, 2019.

[2] M. Buduroh and T. Pudjiastuti, "Colophon in the hikayat pandawa manuscript," *Cultural Dynamics in a Globalized World* , pp. 517–521, 2017.

[3] A. Asi, A. Abdalhaleem, D. Fecker, V. Märgner, and J. El-Sana, "On writer identification for arabic historical manuscripts," *International Journal on Document Analysis and Recognition (IJDAR)*, vol. 20, no. 3, pp. 173–187, 2017.

[4] F. Cloppet, V. Églin, V. C. Kieu, D. Stutzmann, and N. Vincent, "Icfhr2016 competition on the classification of medieval handwritings in latin script," in *2016 15th International Conference on Frontiers in Handwriting Recognition (ICFHR)*, Oct 2016, pp. 590–595.

[5] F. Cloppet, V. Eglin, M. Helias-Baron, C. Kieu, N. Vincent, and D. Stutzmann, "Icdar2017 competition on the classification of medieval handwritings in latin script," in *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, vol. 01, Nov 2017, pp. 1371–1376.

[6] V. Christlein, A. Nicolaou, M. Seuret, D. Stutzmann, and A. Maier, "ICDAR 2019 competition on image retrieval for historical handwritten documents," *arXiv [cs.CV]*, 2019.

[7] Y.-J. Xiong, Y. Wen, P. S. Wang, and Y. Lu, "Text-independent writer identification using sift descriptor and contour-directional feature," in *2015 13th International Conference on Document Analysis and Recognition (ICDAR)*. IEEE, 2015, pp. 91–95.

[8] S. Lai and L. Jin, "Offline writer identification based on the path signature feature," 2019.

[9] H. T. Nguyen, C. T. Nguyen, T. Ino, B. Indurkhya, and M. Nakagawa, "Text-independent writer identification using convolutional neural network," *Pattern Recognition Letters* , vol. 121, pp. 104–112, 2019.

[10] S. Fiel and R. Sablatnig, "Writer identification and retrieval using a convolutional neural network," *Computer Analysis of Images and Patterns*, pp. 26–37, 2015.

[11] S. He, P. Samara, J. Burgers, and L. Schomaker, "Historical document dating using unsupervised attribute learning," in *2016 12th IAPR Workshop on Document Analysis Systems (DAS)*. IEEE, 2016, pp. 36–41.

[12] V. Christlein, M. Gropp, S. Fiel, and A. Maier, "Unsupervised feature learning for writer identification and writer retrieval," *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, 2017.

[13] V. Christlein, D. Bernecker, F. Hönig, A. Maier, and E. Angelopoulou, "Writer identification using GMM supervectors and Exemplar-SVMs," *Pattern Recognition*, vol. 63, pp. 258–267, 2017.

[14] S. Chen, Y. Wang, C.-T. Lin, W. Ding, and Z. Cao, "Semi-supervised feature learning for improving writer identification," *Information Sciences*, vol. 482, pp. 156–170, 2019.

[15] H. T. Nguyen, C. T. Nguyen, T. Ino, B. Indurkhya, and M. Nakagawa, "Text-independent writer identification using convolutional neural network," *Pattern Recognition Letters*, vol. 121, pp. 104–112, 2019.

[16] V. Christlein and A. Maier, "Encoding CNN activations for writer recognition," *2018 13th IAPR International Workshop on Document Analysis Systems (DAS)*, 2018.

[17] V. Christlein, D. Bernecker, F. Honig, and E. Angelopoulou, "Writer identification and verification using GMM supervectors," *IEEE Winter Conference on Applications of Computer Vision*, 2014.

[18] T. Malisiewicz, A. Gupta, and A. A. Efros, "Ensemble of exemplar-SVMs for object detection and beyond," *2011 International Conference on Computer Vision*, 2011.

[19] A. Nicolaou, A. D. Bagdanov, M. Liwicki, and D. Karatzas, "Sparse radial sampling LBP for writer identification," *2015 13th International Conference on Document Analysis and Recognition (ICDAR)*, 2015.

[20] R. Li, W. Liu, L. Yang, S. Sun, W. Hu, F. Zhang, and W. Li, "Deepunet: A deep fully convolutional network for pixel-level sea-land segmentation," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 11, no. 11, pp. 3954–3962, 2018.