



Editorial for special issue on “Advanced Topics in Document Analysis and Recognition”

Josep Lladós¹ · Daniel Lopresti² · Seiichi Uchida³

Published online: 10 August 2021

© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2021

It gives us great pleasure to introduce this special issue focused on state-of-the-art advances in the field of document analysis and recognition. With the world experiencing tremendous disruption due to the COVID-19 pandemic, it is especially impressive to review the quality of the research being conducted in the face of such challenges. We congratulate the authors who contributed successful submissions and thank the reviewers who worked hard on a tight timeframe.

As a result of the open call for papers which was widely disseminated, we received 29 submissions which were judged to be within scope. We encouraged contributions on any topic under the broad umbrella of IJDAR. Each submission was assigned to one of us serving as the guest editors, taking care to avoid any potential conflicts of interest. We solicited reviews from experts in the field following the standard practices for the journal. After a rigorous reviewing process which extended to two or three rounds in some cases, we ultimately accepted 9 papers for publication in this special issue. These reflect both the range of the research in the field today and also the depth of the problems that are being studied.

Several of the papers in this issue address technical challenges that arise when processing historic document images across a variety of time periods and languages. Some address basic steps in the document analysis pipeline (e.g., locating and segmenting text lines), while others attempt to extract higher level information from difficult document images. One paper addresses problems in online handwriting recognition. Many employ various forms of deep learning as an approach that is proving to be powerful and adaptable for document analysis, while one paper, on the other hand, eschews machine learning for reasons of computational efficiency. All reflect the state of the art.

We now turn to providing a brief overview for each paper included in this special issue.

Martin Holeček proposes an integrated system for information extraction from documents in his article, “Learning from similarity and information extraction from structured documents.” Specifically, the author formulates the information extraction task as a 35-class word classification task, and each word class represents, for example, an information tag in invoice, such as “total amount.” For the classification task, the system uses various contextual cues, such as the word-box location, the word recognition result, and the whole document image. For integrating and encoding those cues, a graph convolutional neural network and a transformer are combined. Finally, using about 25,000 document PDFs and three different evaluation scenarios (including “query answer”), the author confirms that the proposed system could achieve higher performance than baselines.

The article “Learning-free Pattern Detection for Manuscript Research: An Efficient Approach Toward Making Manuscript Images Searchable” by Hussein Mohammed, Volker Märgner, and Giovanni Ciotti presents a pattern spotting method where a keypoint feature detection method is combined with the Naive Bayes nearest-neighbor method. This combination realizes a reliable part-based spotting framework without any data-hungry machine learning module. The proposed method has been applied to two datasets,

✉ Daniel Lopresti
lopresti@cse.lehigh.edu

Josep Lladós
josep@cvc.uab.es

Seiichi Uchida
uchida@ait.kyushu-u.ac.jp

¹ Computer Vision Center, Universitat Autònoma de Barcelona, Edifici O, campus UAB, 08193 Bellaterra, Spain

² Department of Computer Science and Engineering, Lehigh University, 113 Research Drive, Bethlehem, PA 18015, USA

³ Department of Advanced Information Technology, Faculty of Information Science and Electrical Engineering, Kyushu University, Kyushu University, 744 Motoooka Nishi-ku, Fukuoka 819-0395, Japan

a dataset of handwriting images on palm leaves and a dataset of medieval manuscript images, and could find similar handwritten patterns and symbols in different images.

In the article “Revealing a History: Palimpsest Text Separation with Generative Networks,” Anna Starynska, David Messinger, and Yu Kong propose a layered-text separation technique for palimpsest manuscripts, which have a three-layer structure by overtext, undertext, and background. Given an overtext image, an input image is reconstructed by estimating undertext and background images by a generative adversarial network (GAN) framework and then merging the overtext image with them. The authors have applied the proposed technique to Archimedes palimpsest images after confirming the expected performance of the proposed technique on an MNIST-based artificial dataset.

The article “A two-step framework for text line segmentation in historical Arabic and Latin document images” by Olfa Mechi, Maroua Mehri, Rolf Ingold, and Najoua Essoukri Ben Amara presents a two-step method for text line segmentation in historical Arabic or Latin document images. First, a deep fully convolutional networks (FCN) architecture is used to segment the text core area. The second step refines the FCN results. It is based on topological structure analysis to extract complete text lines (including the ascender and descender components). Quantitative and qualitative evaluations are reported on a large number of Arabic and Latin document images collected from the Tunisian national archives (ANT) as well as other benchmark datasets.

The article “Self-Supervised Deep Metric Learning for ancient papyrus fragments retrieval” by Antoine Pirrone, Marie Beurton-Aimar, and Nicholas Journet presents a deep metric learning method for the reconstruction ancient papyri with the association of their fragments. The proposed approach uses Deep Convolutional Siamese Networks to explore ways in which papyrologists can obtain useful matching suggestion on new data. The article experimentally shows that the proposed self-supervised approach performs better than a domain transfer method from a large dataset.

“Asking Questions on Handwritten Document Collections” by Minesh Mathew, Lluís Gómez, Dimosthenis Karatzas, and CJ Jawahar addresses the intriguing challenge of adapting Visual Question Answering (VQA) to collections of handwritten documents where full-scale recognition is not yet feasible, but where the answer to a user’s query can be conveyed via selected snippets of images from the collection. They employ a deep embedding network to project both query words and word images into a common high-dimensional space which is then reduced using Principal Components Analysis (PCA). Test results are reported for two specially prepared datasets adapted from existing standard datasets: HW-SQuAD which is a synthetic, handwritten counterpart derived from

SQuAD and BenthamQA which is a smaller set of real scanned manuscript pages selected from the Bentham Collection and annotated for the QA task under study. A comparison is also provided between the authors’ image-based approach and an attempt to use optical character recognition (OCR) on the noisy handwritten inputs.

In the paper “EAML: Ensemble Self-Attention-based Mutual Learning Network for Document Image Classification,” authors Souhail Bakkali, Zuheng Ming, Mickaël Coustaty, and Marçal Rusinol address the problem of combining visual style with textual features to improve the performance of deep CNNs when classifying documents. They employ a self-attention-based fusion module in an ensemble-trainable network that simultaneously learns discriminating features from the visual and text domains. The effectiveness of their approach is demonstrated through testing on two standard datasets: RVL-CDIP which consists of hundreds of thousands of grayscale document images in 16 different classes and the Tobacco-3482 dataset which consists of thousands of document images in 10 different classes. They also perform cross-dataset testing to illustrate the generalizability of their approach.

Applying a deep learning approach to document layout analysis is the focus of the paper “Beyond Document Object Detection: Instance-Level Segmentation of Complex Layouts” by Sanket Biswas, Pau Riba, Josep Lladós, and Umapada Pal. Motivated by the recent success with object detection in computer vision, the authors note that the problem must be redefined in the case of documents owing to the semantic relationships between the various logical components located on a page. Moreover, this work also takes an instance-based approach, working at the pixel level as opposed to the more traditional bounding box approach used in most other work. They test their ideas with a region-based convolutional neural network (Mask-RCNN) using PubLayNet and the HJDataset of historical Japanese documents with complex layouts.

The article “Data Augmentation using Geometric, Frequency, and Beta Modeling approaches for Improving Multilingual Online Handwriting Recognition” by Yahia Hamdi, Houcine Boubaker, and Adel Alimi deals with the problem of data scarcity for learning purposes in deep learning setups. The article presents four data augmentation strategies to improve the performance of recognition systems using small datasets. The four strategies are, respectively, based on: variations in geometric properties, variations of the trajectory high harmonics of the handwritten strikes, the introduction of random change in the parameters of the beta-elliptic model of the handwritten trajectories, and finally a hybrid strategy that combines the previous ones. The proposed approach has been evaluated in the context of multi-lingual online handwriting recognition tasks using end-to-end CNN architecture. Four databases: ADAB, ALTEC-OnDB, Online

KHATT for Arabic script, and UNIPEN for Latin characters, are used for the evaluation.

We believe the research presented here will provide a valuable resource for those working in the field over the coming years. Once again we thank everyone who contributed to the success of this special issue, both authors and reviewers. We also wish to thank journal staff members Katherine Moretti, Gursimaran Kaur, Priya Verma, and Melissa Fearon for their ongoing support and assistance.

Guest Editors:

Josep LLadós.

Daniel Lopresti.

Seiichi Uchida.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.