



Advances in handwriting recognition

Utkarsh Porwal¹ · Alicia Fornés² · Faisal Shafait³

Published online: 9 November 2022

© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2022

Handwriting recognition is the subfield of pattern recognition and document image analysis and recognition that focuses on the recognition and interpretation of handwritten contents, either textual or graphical (e.g., music scores, mathematical expressions, sketches, diagrams, etc.).

Despite decades of research, the recognition of handwritten documents remains challenging because of the varied handwriting styles and document types, ranging from historical documents to modern ones (i.e., documents from administration, bank, insurance companies, etc.). During the last decade, deep Learning has demonstrated great successes in various applications of document analysis and recognition, including, of course, handwriting recognition. Current models make use of Recurrent Neural Networks, Sequence to Sequence models, Transformers, Generative Adversarial Networks, Graph Neural Networks, just to name a few.

Recently, and given the difficulties in obtaining enough labeled data to train deep learning models (which is specially the case of historical documents), data generation, few/zero-shot learning, self-supervised learning, continuous learning, etc., have also been explored. Such improvements make applications more promising, although many research problems remain, especially when processing heterogeneous documents with great variability across writing styles.

This special issue is part of the International Conference on Frontiers of Handwriting Recognition (ICFHR) 2022 held at Hyderabad, India, from December 04–07. It is aimed to report the latest advances in handwriting recognition using

advanced techniques. The guest editors received twelve full submissions, out of which ten are accepted for this special issue. The topics ranged from offline and online handwriting recognition, manuscript alignment to optical music recognition. In addition, several papers presented new datasets for different handwriting analysis tasks. This issue also received one survey paper for historical document image datasets. Each paper went through a peer-review process and received at least two reviews with majority of the papers receiving at least three. All but one paper underwent two rounds of reviews. A summary of all the papers in this issue is as follows:

The article “Offline Handwriting Recognition using Character Spotting” by Nishatul Majid and Elisa H. Barney Smith addresses a problem of unconstrained offline handwritten recognition. Authors expand on their already published idea of offline handwriting recognition using character spotting to make it easily extendable for any script with minor modification. Authors showcase the efficacy of their proposed approach on two different scripts, i.e., Bangla and Hangul/Korean script. Authors also address a major bottleneck in their original approach which required manual annotation of characters in an unsegmented work image. To create the training data for character spotting is a tedious and time-consuming process. Authors also propose a way of automatically tagging character locations in a document. This was a primary contributor in making their approach a generic framework that can work for different alphabetic scripts.

The article “BCBId: First Bangla Comic Dataset and its applications” by Arpita Dutta, Samit Biswas and Amit Kumar Das presents and describes a first Bangla comic book dataset, first of its kind among Indian languages. Comic books and documents are ubiquitous and have been an effective way of storytelling across the world. However, the comic document image research still suffers from lack of enough publicly available datasets, at least in Indian languages. Authors created a Bangla Comic Book Image dataset (BCBId) consisting of over 3300 images from 64 Bangla comic stories written by 8 writers. This dataset has ground truth of different components of comic books, namely panels, characters, speech

✉ Utkarsh Porwal
utkarsh.porwal@gmail.com

Alicia Fornés
afornés@cvc.uab.es

Faisal Shafait
faisal.shafait@seecs.edu.pk

¹ Walmart Inc, 221 River St, Hoboken, NJ, USA

² Computer Vision Center, Autonomous University of Barcelona, Barcelona, Spain

³ National University of Sciences and Technology (NUST), Islamabad, Pakistan

balloons and text lines. In addition to the dataset, authors also present results of sentiment analysis based on the textual content. This dataset will provide a foundation for research in the comic document image analysis for Bangla.

The article “Domain Adaptation for Staff-Region Retrieval of Music Score Images” by Francisco J. Castellanos, Antonio Javier Gallego, Jorge Calvo-Zaragoza and Ichiro Fujinaga addresses a problem of retrieving staff regions in musical scores which is a major step in a broader task of Optical Music Recognition (OMR). Identifying staff regions is crucial as they contain most of the musical information. However, it has largely been posed as a supervised learning problem where first annotation of some manuscripts is needed to train a model. Authors propose a transfer learning-based approach that eliminates the need of manual labeling and leverages existing labeled dataset (source) and adapts the learning from source to a new dataset (target). Authors propose a Domain-Adversarial Neural Network that uses a Gradient Reversal Layer (GRL) to retrieve staff regions within music scores in an unsupervised manner. Authors showcase the efficacy of the proposed approach over a baseline approach where no adaptation is applied and only the source domain is used for training the model.

The article “A holistic approach for image-to-graph: Application to Optical Music Recognition” by Carlos Garrido-Munoz, Antonio Rios-Vila and Jorge Calvo-Zaragoza addresses a problem of Optical Music Recognition (OMR) with a goal of computationally reading music notations from score images. Authors hypothesize that the music notations as graphs where notation primitives such as note heads, flags or stems are considered the nodes of the graphs with syntactic and temporal relationships between these nodes as edges. Authors propose a neural model that takes score images as an input and outputs music notations as graphs. Authors showed the efficacy of their proposed approach on MUSCIMA++ dataset.

The article “A Survey of Historical Document Image Datasets” by Konstantina Nikolaidou, Mathias Seuret, Hamam Mokayed and Marcus Liwicki presents a comprehensive survey of datasets of historical document images, both printed and handwritten. Authors reviewed 56 studies based on different factors ranging from year of publication, number of methods implemented to dataset size and journal outlet. Authors grouped each study into either document classification, layout structure or semantic analysis. Authors provided an overview of the statistics, tasks and benchmarks for datasets used in each study. Authors also discuss the challenges and future direction for these research areas. This is a detailed and systematic study of historical image datasets and different research areas of historic documents.

The article “Combination of explicit segmentation with Seq2Seq recognition for fine analysis of children handwriting” by Omar Krichen, Simon Corbillé, Eric Anquetil,

Nathalie Girard, Elisa Fromont and Pauline Nerdeux addresses children handwriting analysis for French. Authors build on their previous work for children handwriting analysis for the task of copying. In this paper, authors present their approach for analysis for the task of dictation. Authors note that analysis requires accurate segmentation of characters in addition to work recognition because in analysis, errors are highlighted in place. This not only makes segmentation critical for handwriting analysis but it also requires prior knowledge for efficiency. Authors propose using deep learning-based Seq2Seq architecture to encode prior knowledge to drive the analysis process. In addition, authors also combine different prior knowledge strategies to further improve their performance.

The article “A Novel Holistic Unconstrained Handwritten Urdu Recognition System using Convolutional Neural Networks” by Aejaz Farooq Ganai and Farida Khursheed addresses the problem of offline handwritten Urdu recognition. Handwritten Urdu recognition research suffers from a lack of enough publicly available datasets. One of the main datasets, Urdu Nastaliq Handwritten Dataset (UNHD), contains ligatures of only up to five characters, and does not cover the entire Urdu ligature corpus. Authors present a new Urdu Handwritten Ligature Dataset (UHLD) which consists of ligatures of up to seven-character length and covers most of the ligature corpus of the Urdu language. This dataset has 200,000 samples of 2100 basic ligature shapes from 200 different writers. In addition to the dataset, authors also propose an approach for recognition of handwritten Urdu ligatures with up to six-character length. Lastly, this paper also proposes an approach to segment a ligature into characters.

The article “Conv-Transformer Architecture for Unconstrained Off-Line Urdu Handwriting Recognition” by Nauman Riaz, Haziq Arbab, Arooba Maqsood, Khuzaeymah Nasir, Adnan Ul-Hasan and Faisal Shafait also addresses a problem of offline Urdu handwriting recognition. Authors propose a Convolutional Neural Network (CNN)—transformer architecture as a solution. A Convolutional Neural Network (CNN) is used to extract the visual information from the image which is then fed to a full-transformer with three encoder and decoder layers. In addition, authors also use both printed and handwritten text at the time of training to learn more ligatures. Authors showcase the efficacy of their approach on the publicly available NUST-UHWR dataset.

The article “Benchmarking Online Sequence-to-Sequence and Character-based Handwriting Recognition from IMU-Enhanced Pens” by Felix Ott, David Rügamer, Lucas Heublein, Tim Hamann, Jens Barth, Bernd Bischl and Christopher Mutschler addresses a problem of online handwriting recognition (OnHWR). Authors propose several datasets collected using a sensor-enhanced pen for equations and words written in different settings for online seq2seq

tasks. Authors present a dataset for single classification tasks as well. In addition to the data, authors present a broad evaluation benchmark for lexicon-free classification using (Bi)LSTM and Temporal Convolutional Networks (TCN) models combined with CNNs and different Transformer models that can serve as a baseline for future research in the area of OnHWR on paper.

The article “Textline alignment on the image domain” by Boraq Madi, Ahmad Droby and Jihad El-Sana addresses a problem of manuscript alignment which is a critical step in publication of a historic manuscript. Historic manuscripts can have multiple copies available of them from different periods written by different scribes. Therefore, manuscripts are aligned to identify similarities and differences between them. Authors propose an approach that works in image domain to detect subwords, estimate similarity between subwords and to establish alignment between the subword sequences. To showcase the efficacy of their proposed approach, authors also present a new alignment dataset, VML-ALGN, which includes 10 pages from two copies of the same historical Arabic manuscript. The alignment between the two copies is annotated at the subword level on the image domain.

We would like to thank all the authors who contributed to this special issue and express our gratitude to the reviewers who provided constructive and thorough feedback to the authors in a timely manner. Finally, we would like to thank IJDAR Editors-in-Chief—Prof. Koichi Kise, Prof. Daniel Lopresti and Prof. Simone Marinai for their constant support and guidance. Lastly, we would like to recognize the effort of the Springer support staff as this would not have been possible without their help and support.

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