



The 5th TC10/TC11 Summer School on Document Analysis

Monday, 4 July 2023 - Friday, 7 July 2023

Fribourg, Switzerland

Book of Abstracts

Version of June 23

Table of Contents

Reading of Reading for Actuating: Augmenting Human Learning by Experiential Supplements	1
<i>Prof. Koichi Kise</i>	1
You can't hide from tax anymore!, A real world example of how one company has used document analysis and recognition to change the tax industry	2
<i>Rich Kent</i>	2
Deep Learning for Word Spotting: Foundations and Current Developments	3
<i>Prof. Gernot Fink</i>	3
Analysis and understanding of comics: From the detection of basic elements to the creation of semantic links with classic and deep learning-based approaches.	4
<i>Prof. Jean-Christophe Burie</i>	4
Structural methods for document analysis and recognition: From rule-based models to data-driven deep learning	5
<i>Prof. Andreas Fischer</i>	5
Handwriting Recognition in Low Resource Scenarios	6
<i>Prof. Alicia Fornès</i>	6
Large-scale Recognition of Information-rich Documents: From Unreadable Data to Structured Information	7
<i>Prof. Apostolos Antonacopoulos</i>	7

Reading of Reading for Actuating: Augmenting Human Learning by Experiential Supplements

Prof. Koichi Kise

Osaka Prefecture University, Japan

Engaging in reading is a crucial component for language acquisition. By evaluating the reading engagement and behaviors of a learner, we can ascertain the extent of the learner's knowledge and cognitive state. The outcomes of such evaluation can be utilized to enhance learning behaviors through various feedback mechanisms. Generally, the effectiveness of these feedback mechanisms varies among learners and can be identified through the analysis of learning experiences. In this presentation, I will discuss the recent findings from a project titled "Experiential Supplements," with a particular emphasis on its application to language learning. Through the analysis of human experiences, we extract data referred to as experiential supplements, which facilitate the adoption of successful learning strategies employed by other learners. Consequently, learners can acquire a language with heightened efficiency and efficacy.

You can't hide from tax anymore!, A real world example of how one company has used document analysis and recognition to change the tax industry

Rich Kent
CTO of Taina as industrial

The abstract is coming.

Deep Learning for Word Spotting: Foundations and Current Developments

Prof. Gernot Fink

TU Dortmund University, Germany

Machine Learning has made remarkable performance of reading systems possible. Nevertheless, especially with historic documents, there are situations when the complete transcription of a document in question is no longer feasible. This is when approaches for so-called word spotting come into play that today mostly use deep neural networks to build powerful models for document retrieval.

This lecture will first give a brief introduction to the problem of word spotting and the methodological developments in the field. Afterwards, the foundations of current word spotting technology, namely deep neural networks, will be introduced. Then it will be shown how unprecedented retrieval performance can be achieved by adapting convolutional neural networks for the problem of word spotting. Here especially the architecture of the so-called PHOCNet will be covered that has pioneered deep neural models in the field.

However, retrieval performance usually comes at the price that huge amounts of annotated data are required to train high-quality models. This is a major hindrance when thinking about the practical application of word spotting. Therefore, methods will be presented that allow to reduce the required amount of annotated training data significantly by making use of synthetic data, transfer learning and self-training. Another challenge in document retrieval arises from the fact that word spotting normally considers similarity in a purely syntactic fashion. Therefore, methods will be presented that allow to incorporate semantic similarity of search terms in word spotting mode

Analysis and understanding of comics: From the detection of basic elements to the creation of semantic links with classic and deep learning-based approaches.

Prof. Jean-Christophe Burie
University of La Rochelle, France

Born in the 19th century, comics are used for the expression of ideas through sequences of images, often in combination with text and graphics. The comic strip is considered as the 9th art, the sequential art, diffused thanks to the progress of the printing press throughout the world in newspapers, books and magazines. Nowadays, the growing development of new technologies and of the World Wide Web gives rise to new forms of expression fulfilling the paper medium to take advantage of all the freedom of the digital world. However, traditional comics continue to survive and represent an important cultural heritage in many countries.

All the difficulties with the analysis of Comics and Mangas come from the great diversity of the contents. Indeed, each author has his own style both in terms of drawing and text representation. Developing methods to extract and index content from comics is a significant challenge.

In this presentation, I will present the different strategies used to extract the constituent elements of comics such as boxes, speech bubbles, text, characters and the links that may exist between these elements. I will also give an overview of the classical and deep-learning based methods developed to tackle these challenges.

Structural methods for document analysis and recognition: From rule-based models to data-driven deep learning

Prof. Andreas Fischer

University of Fribourg and University of Applied Sciences and Arts Western Switzerland

Documents are rich in structure, especially when considering layouts and tables, but also handwriting, which is composed of individual strokes. In this talk, we will have a closer look at different methods from structural pattern recognition that are well-suited for document analysis, ranging from rule-based models, such as the kinematic theory of rapid human movements and graph edit distance, to data-driven deep learning methods, such as graph convolutional neural networks and transformers. Throughout the talk, I will try to highlight the promising potential of using hybrid approaches, which profit from both rule-based modeling and data-driven learning.

Handwriting Recognition in Low Resource Scenarios

Prof. Alicia Fornès

Universitat Autònoma de Barcelona, Spain

Despite Handwritten Text Recognition (HTR) has been addressed for decades, the recognition of manuscripts is challenging. Current techniques are based on deep-learning architectures, which are powerful but they need a huge amount of training data. Not surprisingly, deep learning methods are not satisfactory in case of low resource scenarios, defined as the scarcity of labeled data for training. This is the case of historical manuscripts, music scores, ciphered manuscripts or documents with rare scripts. This talk will overview the main HTR techniques, with a special focus on methods designed for coping with the limitations of few labelled data, such as few-shot learning, synthetic data generation and self-supervised learning.

Large-scale Recognition of Information-rich Documents: From Unreadable Data to Structured Information

Prof. Apostolos Antonacopoulos
University of Salford, United Kingdom

'A common problem for machine learning, AI and data science systems is the lack of high quality data on which they rely in order to be trained and operate. Even if data is available it is often in a form that cannot be read/operated upon. The talk presents and discusses the issues in the context of the solutions developed by the PRIMa Lab at the University of Salford in a current project delivered for the UK's Office of National Statistics. The objective is to convert large-scale data from an information-rich but unreadable form (census printed volumes) into a highly structured information model that can be used in a variety of applications.