

ICPRAI 2022

3rd International Conference on Pattern Recognition and Artificial Intelligence

Paris
31 May 2022

DOCTORAL CONSORTIUM



Welcome to Doctoral Consortium in ICPRAI 2022

This year, the steering committee of ICPRAI 2022 proposes the first version of the Doctoral Consortium (DC) as a satellite event to the main conference offering the opportunity to PhD students to present their work and meet senior researchers in their field of interest. The Doctoral Consortium at ICPRAI 2022 offers an opportunity for Ph.D. students to test their research ideas, present their current progress and future plans, and to receive constructive criticism and insights related to their future work and career perspectives. For that, a mentor (a senior researcher who is active in the field of ICPRAI topics) was assigned to each participant to provide individual feedback on the student's Ph.D. project. In addition, depending on how they can attend the conference (on-line or on site), students also have the opportunity to present an overview of their research plan during a special poster session (on-site) or an oral presentation (on-line).

The ICPRAI 2022 Doctoral Consortium has accepted 13 students, which have been mentored by 13 senior active researchers in the fields of Artificial Intelligence technics and pattern recognition. During the DC, each research proposal is presented through a teaser/poster session, or an oral presentation, focusing on the outline of the objectives, the methodology, the expected results, the state of the art in their area, and the current stage of their research. During the teaser (introductory) session, each student who is one-site makes a brief presentation of his/her research to the public, inviting to attend the poster session in which the students and their mentors discuss project details. For other on-line PhD students, ten minutes presentations are given and five minutes questions from the audience to explain their objectives and research plan are scheduled.

Welcome to the first DC-ICPRAI 2022 in Paris

Prof. Véronique EGLIN
INSA - Lyon
ICPRAI 2022 Doctoral Consortium Chair

Participants

STUDENTS

Hui Chen
Mathieu Francois
Zhuxian Guo
Adam Harmanec
Amir Khan
Yahjeb Bouha Khattraty
Ana Laura Lezama Sanchez
Magboo Ma. Sheila
Thibault Sauron
Ibrahim Souleiman Mahamoud
Christoph Wehner
Qinghe Zeng

MENTOR

Mounim El Yacoubi
Vincent Poulain d'Andecy
Stéphane Cormier
Rémy Megert
Stéphane Bres
Véronique Eglin
Emanuela Boros
Jean-Christophe Burie
Michaël Coustaty
Bertrand Couasnon
Oriol Ramos Terrades
Giorgio Fumera

Program

14h00-14h10	Introduction / welcome	
14h10-14h20	Teaser slides	
14h20 - 15h50	Oral on-line presentations	
<i>14h20 - 14h35</i>	Hui Chen	Recognition of activities from sensor networks (Internet of Things) based on learning
<i>14h35 - 14h50</i>	Amir Khan	High-Level Inference On-Chip Enabled by Compressive Sensing
<i>14h50 - 15h05</i>	Yahjeb Bouha Khatraty	Thesis topic : Estimation of the quality of rice fields via intelligent systems
<i>15h05 - 15h20</i>	Ana Laura Lezama Sanchez	Método para el descubrimiento de tópicos con aprendizaje profundo
<i>15h20 - 15h35</i>	Ibrahim Souleiman Mahamoud	Automatic and model-free learning of semantic-structural links of fields in a corporate document
<i>15h35 - 15h50</i>	Magboo Ma. Sheila	Using Deep Neural Networks to Perform Brain Tumor Analysis
15h50-16h20	Coffee break	
16h20 - 18h00	Poster session	
	Adam Harmanec	Processing and analysis of super-resolution microscopy images
	Mathieu Francois	Digital transformation of images from technical engineering documents and industrial plans
	Zhuxian Guo	Exploring the tissue and its micro-environment in whole slide imaging to maximize the impact of digital pathology in clinical practices
	Thibault Sauron	Content-based image retrieval applied to radiological image databases
	Christoph Wehner	Interactive and Explainable Link Prediction in Knowledge Graphs
	Ibrahim Souleiman Mahamoud	Automatic and model-free learning of semantic-structural links of fields in a corporate document
18h00-18h15	Conclusion	

Content

Hui Chen, <i>Deep Learning Models for Activity Recognition and Behavior Changes Detection</i> , Université de Sherbrooke, Sherbrooke, Canada	9
Mathieu Francois, <i>Digital transformation of technical engineering documents and industrial plans</i> , Univ. Lyon, INSA Lyon, Villeurbanne, France	11
Zhuxian Guo, <i>Exploring the tissue and its micro-environment in whole slide imaging to maximize the impact of digital pathology in clinical practices</i> , Université Paris Cité, Paris, France	15
Adam Harmanec, <i>Processing and analysis of super-resolution microscopy images</i> , Czech Academy of Sciences, Prague, Czech Republic	17
Amir Khan, <i>High-Level Inference On-Chip Enabled by Compressive Sensing</i> , CSIC-Universidad de Sevilla, Sevilla, Spain	19
Yahjeb Bouha Khattraty, <i>Estimation of the quality of rice fields via intelligent systems</i> , Université Paris 8, Paris, France and University of Nouakchott Al-Aasriya, Nouakchott, Mauritania	21
Ana Laura Lezama Sanchez, <i>Method for Topic Discovery with Deep Learning</i> , Benemérita Universidad Autónoma de Puebla, Puebla, Mexico	23
Thibault Sauron, <i>Content-based image retrieval applied to radiological image databases</i> , Université Paris Cité, Paris, France.....	27
Ibrahim Souleiman Mahamoud, <i>Automatic and model-free learning of semantic-structural links of fields in a corporate document</i> , La Rochelle Université, France.....	31
Christoph Wehner, <i>Explainable and Interactive Machine Learning on Knowledge Graphs for Link Prediction</i> , University of Bamberg, Bamberg, Germany.....	33
Qinghe Zeng, <i>Intelligent Exploration of Whole Slide Image for Digital Histopathology</i> , Université Paris Cité, Paris, France	35

Deep Learning Models for Activity Recognition and Behavior Changes Detection

Hui Chen¹[0000-0003-4401-268X], Charles Gouin-Vallerand¹[0000-0002-5811-2932], Kévin Bouchard²[0000-0002-5227-6602],

Sébastien Gaboury²[0000-0001-7749-3470], Mélanie Couture³[0000-0002-0088-3865], Nathalie Bier⁴[0000-0002-2940-694X], and

Sylvain Giroux¹[0000-0003-0602-5957]

¹ Université de Sherbrooke, Sherbrooke, Canada

² Université du Québec à Chicoutimi, Chicoutimi, Canada

³ CIUSSS West-Central Montréal, Côte-Saint-Luc, Canada

⁴ Université de Montréal, Montréal, Canada

{hui.chen, Charles.Vallerand, Sylvain.Giroux}@usherbrooke.ca,
{Kevin_Bouchard, Sebastien_Gaboury}@uqac.ca,
Melanie.Couture.CVD@ssss.gouv.qc.ca, nathalie.bier@umontreal.ca

Abstract. Deep learning has remarkable contributions to human activity recognition (HAR). Nevertheless, it is still challenging in the long term to recognize the daily activities of the older people in a real smart home with imbalanced and limited labeled noisy sensor data and associate the behavior changes with health status transition. In this research, we propose a semi-supervised deep model to recognize the older people's activities of daily living (ADLs) and convolutional auto-encoder (CAE) to detect behavior changes based on massive ambient sensors data in smart homes.

Keywords: Human Activity Recognition (HAR), Deep Learning, Activities of Daily Living (ADLs).

1 Introduction

Deep learning (DL) models are recently widely used for sensor-based activity recognition [1]. Compared to traditional pattern recognition methods, deep learning could perform automatic high-level feature extractions to perform better without relying on human-crafted feature extraction. DL models could achieve competitive performance by monitoring human activities and physiological parameters to detect the abnormal condition early, and unusual behaviors [2]. However, there are still many challenges remaining. It is time-consuming and costly to manually label sensor data to get the ground truth in a real environment for a long period (e.g., several months). The sensor data collected in real smart homes often include noisy information and imbalanced activities that impact activity recognition accuracy. It is challenging to evaluate behavior changes associated with health status changes or functional changes.

My Ph.D. thesis aims to provide DL models to recognize daily life activities (e.g.: eating and sleeping) using partially labeled ambient sensor data collected in real smart homes and detect older people's behavior changes in visualized routine profiles.

2 Proposed Methods

We propose two deep learning models to tackle the above research problems. First, a semi-supervised deep learning model recognizes human activities with limited labeled, imbalanced, and noisy real ambient sensor data. Second, we use CAE to detect behavior changes using visualized daily activities density maps [3].

2.1 Activity Recognition with Partially Labeled Sensor Data

With the inspiration of [4][5] for noisy information and imbalanced activities, we plan to use a semi-supervised deep model to recognize partially labeled multimodal unbalanced real sensor data. First, we run deep models (e.g., CNN, LSTM, and Recurrent neural network (RNN)) iteratively on the small amount of labeled data and minimize the classification errors. Second, these models predict the unlabeled data. Then choose the most confident "*pseudo-labels*" using majority voting and label unlabeled sets based on the class rebalancing scheme for subsequent iterations. Finally, the ensemble models re-train both "*pseudo-labeled*" data and the original labeled data. This procedure is iterated until no more unlabeled data remains.

We would evaluate this model on the dataset based on an older people living community called the SAPA project [6]. Around thirty smart homes have been installed with the internet of things (IoT) infrastructure, such as passive infrared

(PIR) motion, contact, and door/window sensors. The sensor events were continuously collected for more than two years.

2.2 Detect Behavior Changes

We plan to use two steps to solve the challenge of evaluating behavior changes associated with health status changes.

First, we use a cluster-based model (e.g., K-means) to group activities, time, and locations to daily life patterns on long-term sensors data. The personalized routine profile is generated using these patterns and older people's health profiles (e.g., memory problems and medication status) provided by clinicians.

Second, we propose CAE to detect long-term behavioral changes (e.g., new behaviors and changes in the duration of activities). With the inspiration of [3][7], we first generate visualized activity density map based on the period window size (e.g., three months or six months) using recognized daily activities. Then we plan to use the CAE structure to detect behavioral changes in these activity density images. The encoder of CAE included two convolutional layers with a kernel size of five and three, followed by a fully connected layer. The decoder was a mirror of the encoder with convolutional transpose layers. Finally, we would calculate latent features between different period windows using a distance-based anomaly score to detect the long-term behavioral changes.

The clinicians' research notes of health status changes (e.g., hospitalization) and the feedback of healthcare providers and residents would aid us in evaluating and associating the abnormal behavior changes with health status changes.

3 Research Plan, Mentor Exchanges, and Expected Contributions

I started the Ph.D. research in November 2020 and would finish in December 2023. I am currently developing the labeling application for labeling the raw sensor data and recognizing daily activities using semi-supervised deep models (e.g., CNN and LSTM). I plan to deploy the application in real smart homes for future work. Older people are encouraged to mark daily activities using labeling applications to provide ground truth. Then I would recognize daily activities and detect behavior changes using proposed models on real smart homes' sensor data.

Thanks to my mentor Prof. Mounim A. El-Yacoubi, who provided valuable suggestions about my research. (1) Use a weakly supervised learning model to leverage the last two years' unlabeled sensors data for activity recognition. (2) Refer to anomaly detection to improve the CAE model to detect the long-term behavior changes using a period window. (3) How to clarify and evaluate the long-term health changes of the seniors using the proposed models.

The expected contributions of this research are to use DL models to recognize human activities and detect behavior changes in the real world which is one of the main challenges in the HAR community. The expected results of early change detection also contribute to the home care of the aging people community to improve home care and reduce health care costs.

References

1. Jindong Wang, Yiqiang Chen, Shuji Hao, Xiaohui Peng, and Lisha Hu: Deep learning for sensor-based activity recognition: A survey. *Pattern Recognition Letters*, 119:3–11, 2019.
2. Giovanni Diraco, Alessandro Leone, Andrea Caroppo, and Pietro Siciliano: Deep learning and machine learning techniques for change detection in behavior monitoring. pages 38–50. *CEUR Workshop Proceedings*, 2020.
3. Shuang Wang, Marjorie Skubic, and Yingnan Zhu: Activity density map visualization and dissimilarity comparison for eldercare monitoring. *IEEE Transactions on Information Technology in Biomedicine*, 16(4):607–614, 2012.
4. Chen Wei, Kihyuk Sohn, Clayton Mellina, Alan Yuille, and Fan Yang: Crest: A class-rebalancing self-training framework for imbalanced semi-supervised learning. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 28:10852–10861, 2021.
5. Kaixuan Chen, Lina Yao, Dalin Zhang, Xianzhi Wang, Xiaojun Chang, and Feiping Nie: A Semisupervised Recurrent Convolutional Attention Model for Human Activity Recognition. *IEEE Transactions on Neural Networks and Learning Systems*, 31(5):1747–1756, 2020.
6. Hubert Ngankam., Maxime Lussier., Aline Aboujaoudé, Cédric Demongivert, Hélène Pigot, Sébastien Gaboury, Kevin Bouchard, Mélanie Couture, Nathalie Bier, and Sylvain Giroux: Sapa technology: An aal architecture for telemonitoring. pages 892–898. *SciTePress*, 2022.

Doctoral Consortium - ICPRAI 2022

Thesis title : Digital transformation of technical engineering documents and industrial plans

PhD Student : Mathieu Francois, Thesis supervisor : Véronique Eglin and Mentor : Vincent Poulain d'Andecy

Univ Lyon, INSA Lyon, CNRS, UCBL, LIRIS, UMR5205, F-69621 Villeurbanne, France
francois.mathieu@orinox.com

CIFRE Contract

Estimated start and end of the thesis : 05/2021 - 05/2024

1 Summary of the thesis topic

Today, a large number of engineering documents are stored either in paper format (printed version of a native digitized document or single representation medium), or in a weakly structured digital representation (pdf documents), or in their native digitized format (as images). The digital transformation of these contents is an ambitious research topic in the industrial world. Companies involved in this digital transformation hope to access digital resources in an automatic, indexed and enriched way (with for example new semantic relations between digitized contents, the search for information in indexed and annotated datasets, the intelligent edition of schemas and P&ID).

Engineering documents or other technical documentation of industrial installations are composed of a large number of elements of different nature, see figure 2. There are for example texts with or without semantics, graphic symbols and drawings... These documents can be very dense with information, which can even overlap. For a human with engineering knowledge, it is quite obvious to distinguish this information but what about a machine?

The automatic interpretation of these engineering plans can be very complex for several reasons. First of all, these documents are very poorly (if at all) structured. The information (text/symbol) can be represented in any orientation, size, color... The information on these plans is often very dense and can overlap, making it difficult to understand.

For the edition of these documents, the experts do not have real standards allowing to draw the plans. This documentation is made up of very large digitized technical plans (type A0) whose graphic, symbolic and textual content is heterogeneous, of variable quality and respects representation standards that can vary from one document to another. For example, the same technical component can be represented symbolically in several ways depending on the standard used.

Concerning the thesis work, the first topic of interest were the extraction of textual entities within these documents and then, the extraction of symbols. These parts were the first in order to be able, in the future, to link and interpret the two types of information in order to correct possible prediction errors and make sense of them.

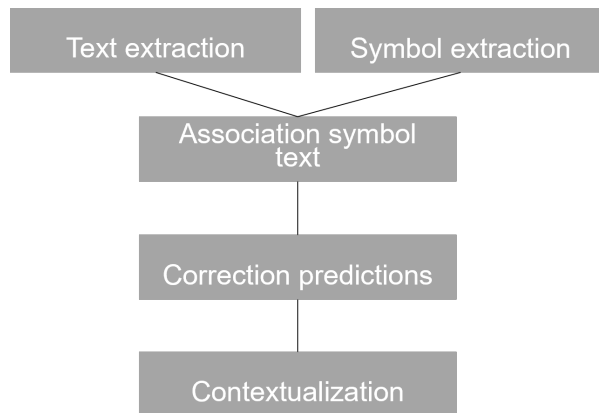


Fig. 1. Diagram representing the different stages of the thesis.

2 Realization

2.1 Work on text detection and recognition on engineering documents

During the first year, we were interested in the detection and recognition of printed text (short texts of variable orientations with specific business semantics). The strings present on this type of document are thus very little structured (few sentences, few nominal groups but texts named tags which qualify the graphic regions of the plans). The textual entities can thus be placed anywhere on the document, they can be of any orientation, size or color.

The first phase of this work on text extraction was to make a state of the art of already existing technologies that could answer our problems. In order to carry out this bibliography, the subject have been separated into 4 main parts: image pre-processing, text detection, text recognition and error correction.

In order to build a complete detection and recognition model, this part of the thesis has been started by creating a dataset from a selection of plans and maps of the Orinox company. An annotation tool has been developed to extract the coordinates of the text areas on the PDF documents and to include the corresponding transcripts.

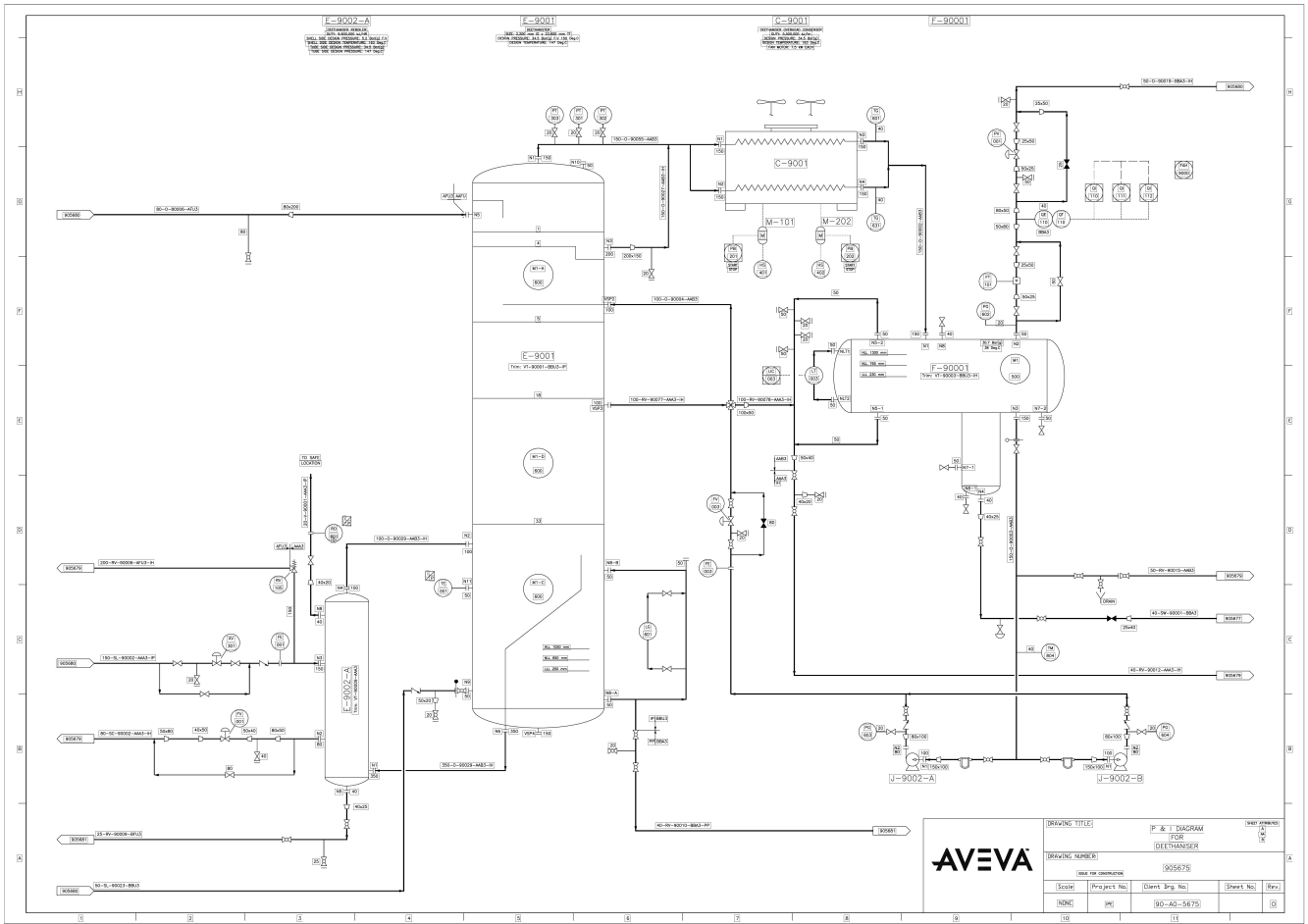


Fig. 2. Example of an engineering document containing multi-oriented text and graphic symbols.

After that, some technologies identified in the state of the art [1][3][4] has been tested . Then we started the development of a text detection solution based on the EAST algorithm [1] to verify that FCN technologies, which are generally used in natural scene contexts, are adaptable to our use case. The EAST algorithm is a fully convolutional network (FCN) model that proposes predictions of the word or line containing the text, excluding intermediate steps such as candidate proposal, text region formation and word partitioning. The generated text predictions are sent to non-maximal suppression to obtain the final results.

Then, an open-source OCR has been used to recognize the previously detected characters. This one is called Tesseract and it is based on a neural network architecture of the LSTM type. It is configured in order to each image

given as input corresponds to a single line of text (corresponding to the output of our detection system). We chose this OCR because it is open-source and easy to use. There are better OCRs on the market but our problem was not pure character recognition.

Among the texts present in the documents, some strings of characters called Tags are of particular interest to us (see Fig. 3).

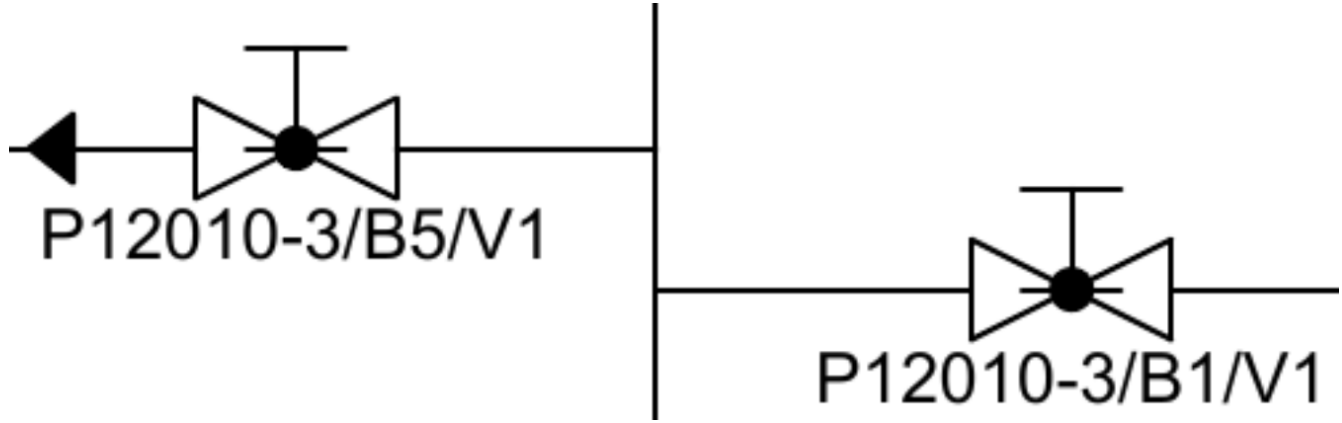


Fig. 3. Example of tags with their associated symbol.

Tags are identifiers of technical elements present on the documents. They are not linked to any lexicon or language model. They are usually composed of a capital letter, a number and a separator character (example: AA-1234-BB). Because of these properties, they are frequently the target of transcription errors (even when dissociating the dictionary from the transcriber).

In order to regain control of the transcription errors, we turned to a system allowing the Post-OCR correction of these Tags. This one is based on a clustering (Affinity Propagation[2]) and a string structure analysis. The Affinity Propagation algorithm has the ability, compared to other clustering methods, to not need the number of input clusters. To keep it simple, the data communicate with each other to measure their level of similarity and then the clusters are formed.

This first work realized during the first year of the thesis was the subject of a long article (15 pages) in the international conference DAS (Document Analysis System 2022, La Rochelle) and will be presented on May 24, 2022.

In this study, a F1 Score of 84% for the text detection system has been obtained. Some remaining errors are due to the wrong understanding of the symbols (symbol confused with text, enlargement of the text zones wrongly...). That's why decided to put this part on pause and to resume it when the symbolic entities extraction system will be operational.

Regarding the post-OCR correction, a gain of 7 % was obtained by calculating the difference between the WER of the OCR output and the post-OCR correction. We are sure that the results can be improved with more data. Indeed, this module has been processed shot by shot. But if we imagine to have several data from the same project, we could apply the clustering method on several plans, which would give a better representation of the clusters.

2.2 Work on symbol detection and recognition on engineering documents

Tags are by nature difficult texts to recognize because their semantics is deduced from the construction of their sequences of characters and digits. In order to avoid confusion in the recognition of these sequences, we have chosen to couple the recognition of the text and the neighboring graph: the association of the two information modalities seems to be an interesting track never explored before in the context of industrial maps and plans.

For this purpose, we are currently performing a first state of the art on the detection of objects close to the graphical objects of the enterprise and thinking about a model of coupling text-graphical modalities to jointly improve the transcription/post-correction of texts and symbols, see figure 4.

As for the work concerning textual entities, we need annotated data to train the detection models. Indeed, engineering documents are specific and often private data: it is very difficult to find a public dataset allowing to meet this training challenge with the required level of annotation.

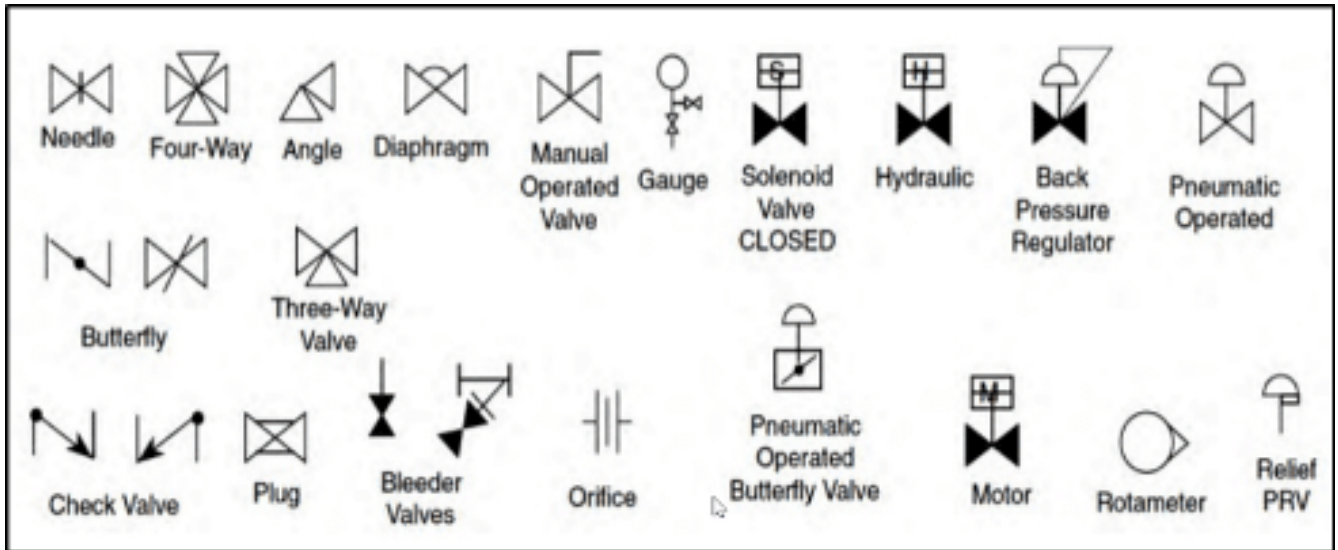


Fig. 4. Example of engineering symbols.

As for the first part of the thesis concerning the tags and texts of nomenclatures, we have created a second annotation tool allowing to extract information from engineering documents directly in the drawing software of these documents. This tool has the task of locating the symbols and providing the coordinates of the bounding boxes. We then named and classified the symbolic data.

This work of detection and text-graphic coupling is a complex task that we are currently starting. To start this step, we started from state of the art technologies [5][6][7][8] to identify the technological locks to our use case and to advance in our proposals.

3 Perspectives for the thesis year 2

The objectives for the work during the second year of the thesis are as follows:

- Continue work on symbol detection and recognition
- Focus on the detection of links and the coupling between the different information present on the documents (text and graphic symbols)
- Publish on the occasion of the ICDAR 2023 conference

References

1. X. Zhou et al., "EAST: An Efficient and Accurate Scene Text Detector," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2642-2651, (2017).
2. D. Dueck, "Affinity propagation: clustering data by passing messages," Ph.D. dissertation, Citeseer, (2009).
3. Tian, Z., Huang, W., He, T., He, P., & Qiao, Y. (2016). Detecting Text in Natural Image with Connectionist Text Proposal Network. ECCV.
4. Linjie Deng, Yanxiang Gong, Yi Lin, Jingwen Shuai, Xiaoguang Tu, Yuefei Zhang, Zheng Ma, Mei Xie, Detecting multi-oriented text with corner-based region proposals, Neurocomputing, Volume 334, 2019, Pages 134-142, ISSN 0925-2312.
5. K. He, G. Gkioxari, P. Dollár and R. Girshick, Mask R-CNN, 2017 IEEE International Conference on Computer Vision (ICCV), 2017, pp. 2980-2988.
6. J. Redmon and A. Farhadi, "YOLO9000: Better, Faster, Stronger," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 6517-6525.
7. J. Long, E. Shelhamer and T. Darrell, "Fully convolutional networks for semantic segmentation," 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 3431-3440.
8. M. Luqman, T. Brouard and J. Ramel, "Graphic Symbol Recognition Using Graph Based Signature and Bayesian Network Classifier," in 2013 12th International Conference on Document Analysis and Recognition, Barcelona, Spain, 2009 pp. 1325-1329.

PhD Student: Zhuxian GUO

Laboratory of Informatics Paris Descartes (LIPADE), Paris 75006, France

1. Title of the thesis

Exploring the tissue and its micro-environment in whole slide imaging to maximize the impact of digital pathology in clinical practices.

2. Starting and expected finalization date of the PhD

Starting date: November 1, 2021

Expected finalization date: October 31, 2024

3. Supervisor of the thesis

Paris: Nicolas LOMENIE / Camille KURT Switzerland: Henning Muller

4. context and objectives of the PhD

We have two ongoing projects with Ambroise-Paré Hospital (POCHI project) and Henri- Mondor Hospital (MONDOR project), respectively. We also have one upcoming national project: Translational Cancer Research Program (PRT-K).

In the POCHI project, we are conducting a phase II clinical trial on linear quantification of lymphoid infiltration of the tumor margin in colorectal carcinoma (CRC). The study on linear quantification of CD3 infiltration within a particular region across the invasive front helps on therapy outcome forecasting for CRC patients [1].

In the MONDOR project, we are exploring machine learning approaches to tackle the challenges in transcriptomic subclass distinction based on Whole Slide Images (WSI) of hepatocellular carcinoma (HCC), which is a highly molecular and histological heterogeneous cancer. HCC phenotype appears to be closely related to particular gene mutations [5], [8], [6]. Applying clustering on the morphological appearance of the digitalized histopathological specimens (i.e., WSI) is helpful in HCC transcriptomic subclass distinction with the absence of mRNA sequencing, which is expensive and is not a routine examination in clinical practices. The study can help with HCC diagnosis, prognostication, novel biomarker discovery and personalized medicine development.

PRT-K will use several prospective cohorts of stage II and III colon cancer, involving more than 3600 patients. We will generate and validate a prognostic algorithm using machine learning cutting-edge technology to estimate a patient's survival using a routine Hematoxylin Eosin Safran (HES) WSI. The final objective of this program is to generate an open-source free software to determine patient prognosis based on a unique digitized HES slide.

5. Summary of the progress made so far

Due to the inherent deviation of the annotations, it will be interesting to explore some recently available semi-supervised or unsupervised models (e.g., MixMatch[2], RUC [10], SCAN [11]) and the possible application of models in semi-supervised/unsupervised fashion in digital pathology (e.g., [9], [3]). Recently, Some weakly supervised models (e.g., [7]), which only require slide-level annotation, are also popular in the pathology community.

[4] pointed out the limitations of CNNs in digital pathology applications. There are limitations such as relatively small field of view compared to a large-scale gigapixel histopathological slides and the incapability of CNNs in understanding non-Euclidean information. Graph neural networks (GNNs), which have been used in non-Euclidean data mining and machine learning fields, can be promising in computational pathology (e.g., [12]).

Now I am working on the dataset from TIGER: <https://tiger.grand-challenge.org/> and from PAIP 2019 (<https://paip2019.grand-challenge.org/>)

6. Future research plan

Phase 1: Studying state-of-the-art CNN-like architectures for digital pathology and for other applications in medical image processing. Studying the fresh supervised and unsupervised deep models in computer vision and thinking about their applications in our projects related to CRC and HCC. Contributing to the ongoing and upcoming projects.

Phase 2: To develop a robust and reproducible architecture to recognize different morphological appearance of the main HCC phenotypes. Clustering the visual representations, extracted by the architecture, of slides with certain phenotypes to particular gene mutations by learning the WSI and their corresponding mRNA sequencing data. Also, to find a way to perform a single analysis of a larger field of vision on a slide, which is difficult due to the limitation of patch size feeding to a CNN, and to learn the interactions between different

architectures in this field. Finally, improving the interpretability of the whole tissue analysis system to help people intuitively understand the underlying pathological mechanisms (e.g., visualization), and to improve the acceptance and application scope of the pathology community for the tissue analysis using neural networks. Phase 3: Exploring the possibility of generalizing the system developed in Phase 2 to slides of other tissues, slides with other staining protocols (e.g., Immunohistochemistry Stain), etc. Exploring the introduction of augmented reality (AR) technology to conventional microscopes and to other pathology laboratory equipment to assist pathologists in examining tissue specimens. Publications on the journals/conferences in the digital pathology community and submitting a doctoral dissertation.

7. References

- [1] Allard, M.-A., Bachet, J. B., Beauchet, A., Julie, C., Malafosse, R., Penna, C., Nordlinger, B., and Emile, J.-F. Linear quantification of lymphoid infiltration of the tumor margin: a reproducible method, developed with colorectal cancer tissues, for assessing a highly variable prognostic factor. *Diagnostic Pathology* 7, 1 (11 2012), 156.
- [2] Berthelot, D., Carlini, N., Goodfellow, I., Papernot, N., Oliver, A., and Raffel, C. MixMatch: A Holistic Approach to Semi-Supervised Learning. *arXiv* (5 2019).
- [3] Hatami, N., Bilal, M., and Rajpoot, N. Deep Multi-Resolution Dictionary Learning for Histopathology Image Analysis. *IEEE TRANSACTIONS ON MEDICAL IMAGING* XX (4 2021), 1.
- [4] He, Y., Zhao, H., and Wong, S. T. Deep learning powers cancer diagnosis in digital pathology. *Computerized Medical Imaging and Graphics* 88 (3 2021), 101820.
- [5] J, C., G, C., S, I., G, A., E, L., JF, B., C, L., Y, H., D, A., P, B.-S., JC, N., and J, Z.-R. Histological subtypes of hepatocellular carcinoma are related to gene mutations and molecular tumour classification. *Journal of hepatology* 67, 4 (10 2017), 727–738.
- [6] J, C., M, Z., V, P., and J, Z.-R. Molecular and histological correlations in liver cancer. *Journal of hepatology* 71, 3 (9 2019), 616–630.
- [7] Lu, M. Y., Williamson, D. F. K., Chen, T. Y., Chen, R. J., Barbieri, M., and Mahmood, F. Data Efficient and Weakly Supervised Computational Pathology on Whole Slide Images. *Tech. rep.*, 2020.
- [8] M, Z., N, P., G, A., A, L., JC, N., F, O., C, C., S, M., M, B., C, F., GP, P., J, R., T, D., A, L., B, G., V, V., C, A., J, D., C, C., J, Z.-R., N, B., O, S., N, G.-C., V, P., and J, C. Macrotrabecular-massive hepatocellular carcinoma: A distinctive histological subtype with clinical relevance. *Hepatology (Baltimore, Md.)* 68, 1 (7 2018), 103–112.
- [9] Muhammad, H., Sigel, C. S., Campanella, G., Boerner, T., Pak, L. M., B'uttner, S., IJzermans, J. N. M., Koerkamp, B. G., Doukas, M., Jarnagin, W. R., Simpson, A., and Fuchs, T. J. Towards Unsupervised Cancer Subtyping: Predicting Prognosis Using A Histologic Visual Dictionary. *arXiv* (3 2019).
- [10] Park, S., Han, S., Kim, S., Kim, D., Park, S., Hong, S., and Cha, M. Improving Unsupervised Image Clustering With Robust Learning.
- [11] Van Gansbeke, W., Vandenhende, S., Georgoulis, S., Proesmans, M., and Van Gool, L. SCAN: Learning to Classify Images without Labels. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 12355 LNCS (5 2020), 268–285.
- [12] Zhou, Y., Graham, S., Koohbanani, N. A., Shaban, M., Heng, P.-A., and Rajpoot, N. CGC-Net: Cell Graph Convolutional Network for Grading of Colorectal Cancer Histology Images. *Proceedings - 2019 International Conference on Computer Vision Workshop, ICCVW 2019* (9 2019), 388–398.

PROCESSING AND ANALYSIS OF SUPER-RESOLUTION MICROSCOPY IMAGES

ICPRAI 2022 - DOCTORAL CONSORTIUM SUBMISSION

Adam Harmanec^{1,2}, Supervisor: Filip Šroubek¹, Co-supervisor: Zuzana Kadlecová^{2,3}

¹Institute of Information Theory and Automation, Czech Academy of Sciences, Prague, Czech Republic

²Department of Histology and Embryology, Faculty of Medicine, Masaryk University, Brno, Czech Republic

³Cambridge Institute for Medical Research, University of Cambridge, Cambridge, UK

1. RESEARCH PLAN

The PhD research began in October 2021 and is expected to take 4 years, that is until circa October 2025.

1.1. Summary

Microscopy has always been an essential tool for biological research. The resolution of an optical microscope is inherently limited by the finite wavelength of light [1]. Over the last two decades, numerous super-resolution imaging techniques have been invented to surpass this limit. This super-resolution revolution has transformed the way cells and tissues are studied. This project will focus on methods and data obtained from Total internal reflection fluorescence - structured illumination microscopy (TIRF-SIM) [2, 3]. The high spatial and temporal resolution and the low illumination power of TIRF-SIM are optimal for live-cell imaging which is challenging with other super-resolution techniques. However, this new data represents a significant challenge for data analysis, and there are currently no objective and automatic methods. Hence, to bridge this gap, this project aims to develop novel image processing approaches suitable for this type of images and videos and provide the biomedical community with an essential tool for unbiased analysis of experiments studied with TIRF-SIM. We aim to develop methods and algorithms that are generic and applicable to a wide range of tasks, not only to TIRF-SIM data analysis, but for this project, they will be applied to and validated on data from plasma membrane research in collaboration with the Institute of Medical Research at the University of Cambridge. We assume that the project's first year will be dedicated to data assessment, optimization of the SIM reconstruction algorithm and generation of synthetic training data that could be used to train detection and classification deep machine learning models. The aim of the consequent years is to properly integrate image processing methods and thus create a robust pipeline for automatic and unbiased analysis of TIRF-SIM data.

1.2. Current State of Knowledge

The concept of structured illumination microscopy (SIM)[3] for doubling the resolution of optical microscopes was introduced over two decades ago. It's acquisition speed and low phototoxicity make it one of the best options for observing processes in living cells. It was soon combined with Total internal reflection fluorescence (TIRF)[2, 4], a method for studying very thin layers of samples (<100 nm). To obtain a detailed and noise-free super-resolved image, the illumination parameters must be known precisely. These have to be estimated from the original images, which is challenging when observing living cells because they are obtained at high frame-rates with weak laser illumination. Various methods have been proposed to improve the estimation from images with low signal-to-noise ratio [5].

Additionally, the reconstruction algorithm itself can be modified to better cope with noise [6]. It has recently been shown that deep neural networks can be trained to directly perform SIM reconstruction [7]. However, these published approaches are not instantly adaptable to our use case and necessitate further modification.

Detection, classification and tracking of objects in images are some of the most common tasks in computer vision. In recent years, deep learning has become a very successful and popular solution [8]. However, a lot of the research is concerned with diverse natural images and is not directly transferable to microscopy images, to which only limited attention is paid. To our knowledge, no deep learning models have been applied to the analysis of TIRF-SIM images and videos.

We are unaware of any off-the-shelf tools for unbiased, automatic, and quantitative data analysis from TIRF-SIM microscopy, especially for videos capturing the dynamics of vesicle formation at the plasma membrane. Therefore our research can be inspired but not build on the oversimplified semi-automatic image processing methods used by Willy et al [9] and Zaccai, Kadlecová et al [10], which are based on tools not intended and not very suitable for the purpose.

1.3. Methods of Research

The project’s first step is to assess the quality, robustness and reproducibility of the available TIRF-SIM data and the methodology used for its acquisition. Ideally, this should include a visit to the microscope at the University of Oxford and a consultation with its constructor and operator, Dr. Kseniya Korobchevskaya. We believe the data acquisition process could be tweaked to take into account the processing of its results. Our preliminary data inspection indicated the vital need for a better reconstruction method since the low signal-to-noise ratio and subtle cell migration during acquisition leads to reconstruction artifacts. On the other hand, this time investment in optimizing reconstruction procedures will later pay off in simplified downstream data analysis. To that end, we want to experiment with new Wiener regularizations and other deconvolution techniques.

The second step is to develop methods for object detection and classification for the resulting super-resolved images. Algorithms used until now were optimized for the lower resolution and cannot be applied to this new data. The observed cell structures highlighted by fluorescent markers should be found in each frame and segmented as accurately as possible. We propose to use deep convolutional neural networks adapted to the specific properties of the subjects, such as their large number but the relatively coherent size and shape. We also want to experiment with using the images in their original resolution as input to the network, which could eliminate the residual effects of the reconstruction at the cost of a more complicated training procedure. Deep learning notoriously requires a large amount of data which we intend to overcome by using synthetic images.

The third step is to match and track the detected objects in consecutive frames in order to monitor their evolution over time for up to several minutes. This comes with a number of challenges. For example, new objects may appear, and the existing ones may transform and disappear. In addition, the approach should be robust enough to handle the formation and breakup of local clusters of objects and their movement. To this end, we want to repurpose the rich research of cell and particle tracking to construct trajectories from individual detections.

The final step is to process and analyze the obtained tracks. The aim is to provide biological experts with objective and informative statistics on the properties and behavior of the monitored objects. Specific analytical procedures may evolve and change as they are directly motivated by the needs of cellular biological research, but the previous steps should be robust and generic enough to be easily adaptable to a variety of use cases.

We expect to fully complete the first step and start working on the second step during the project’s first year. After that, the next steps will be accomplished sequentially in the following years.

2. REFERENCES

- [1] Ernst Abbe, “Beiträge zur theorie des mikroskops und der mikroskopischen wahrnehmung,” *Archiv für mikroskopische Anatomie*, vol. 9, no. 1, pp. 413–468, 1873.
- [2] Daniel Axelrod, “Cell-substrate contacts illuminated by total internal reflection fluorescence,” *Journal of Cell Biology*, vol. 89, no. 1, pp. 141–145, 1981.
- [3] Mats GL Gustafsson, “Surpassing the lateral resolution limit by a factor of two using structured illumination microscopy,” *Journal of microscopy*, vol. 198, no. 2, pp. 82–87, 2000.
- [4] Reto Fiolka, Markus Beck, and Andreas Stemmer, “Structured illumination in total internal reflection fluorescence microscopy using a spatial light modulator,” *Optics letters*, vol. 33, no. 14, pp. 1629–1631, 2008.
- [5] Junchao Fan, Xiaoshuai Huang, Liuju Li, Shan Tan, and Liangyi Chen, “A protocol for structured illumination microscopy with minimal reconstruction artifacts,” *Biophysics Reports*, vol. 5, no. 2, pp. 80–90, 2019.
- [6] Xiaoshuai Huang, Junchao Fan, Liuju Li, Haosen Liu, Runlong Wu, Yi Wu, Lisi Wei, Heng Mao, Amit Lal, Peng Xi, et al., “Fast, long-term, super-resolution imaging with hessian structured illumination microscopy,” *Nature biotechnology*, vol. 36, no. 5, pp. 451–459, 2018.
- [7] Luhong Jin, Bei Liu, Fenqiang Zhao, Stephen Hahn, Bowei Dong, Ruiyan Song, Timothy C Elston, Yingke Xu, and Klaus M Hahn, “Deep learning enables structured illumination microscopy with low light levels and enhanced speed,” *Nature communications*, vol. 11, no. 1, pp. 1–7, 2020.
- [8] Yuxin Wu, Alexander Kirillov, Francisco Massa, Wan-Yen Lo, and Ross Girshick, “Detectron2,” <https://github.com/facebookresearch/detectron2>, 2019.
- [9] Nathan M Willy, Joshua P Ferguson, Salih Silahli, Cemal Cakez, Farah Hasan, Yan Chen, Min Wu, Henry C Chang, Alex Travesset, Siyu Li, et al., “Endocytic clathrin coats develop curvature at early stages of their formation,” *Biorxiv*, p. 715219, 2019.
- [10] Nathan R Zaccai, Zuzana Kadlecova, Veronica Kane Dickson, Kseniya Korobchevskaya, Jan Kamenicky, Oleksiy Kovtun, Perunthottathu K Umasankar, Antoni G Wrobel, Jonathan GG Kaufman, Sally R Gray, et al., “Fcho controls ap2’s initiating role in endocytosis through a ptdins (4, 5) p2-dependent switch,” *Science advances*, vol. 8, no. 17, pp. eabn2018, 2022.

High-Level Inference On-Chip Enabled by Compressive Sensing^{*}

Amir Khan^{Student}
Jorge Fernández-Berni^{Supervisor} and Ricardo Carmona Galán^{Supervisor}

Instituto de Microelectrónica de Sevilla, CSIC-Universidad de Sevilla, Sevilla, Spain
khan@imse-cnm.csic.es

1 Important Dates

PhD start date: 01-December 2020.

Expected PhD end date: 30-November 2024.

2 Presentation of Context

Compressive Sensing (CS) [2] has attracted researchers from a variety of fields because of its potential for low-power operation, reduced bandwidth, and low memory consumption. CS was originally proposed for concurrent acquisition and compression of sparse signals, and eventual signal reconstruction from a significantly smaller set of samples than those required by the Nyquist-Shannon theorem [3]. However, the reconstruction of a CS signal requires solving an optimization problem, which is normally a computationally intensive task, while keeping a high signal-to-noise ratio. It has been estimated that the reconstruction problem is four orders of magnitude more complex in terms of computation than CS-based acquisition [4]. Reconstruction is not always required though. Significant efforts were made to understand the amount of information preserved in CS samples for inference purposes. It was shown that feature extraction and decision making can be directly conducted in the compressed domain, known as ‘compressed learning’ (CL), with nearly the same accuracy as in the uncompressed domain—also known as data domain [1].

As of now, to the best of our knowledge, there exists no single on-chip system implementation performing inference in the CS domain. In this thesis, we aim to develop an on-chip sensor-processor system to perform inference in the CS domain.

3 Thesis Objective(s)

3.1 General Objective

The generic objective of this thesis is the early extraction of the relevant information contained in the visual stimulus for automatic in-sensor object recognition and scene interpretation. To this end, we are going to use compressive sampling as a fundamental tool, based on the fact that visual information in natural images can be described with very few data if they are expressed in the appropriate basis. Our objective is to implement a CMOS smart sensor-processor chip exploiting CS for on-chip inference based on the hypothesis that working in the compressed domain allows implementing high-level inference under severe restrictions on computational and power resources.

3.2 Specific Objectives

We propose the following specific objectives:

1. Design of machine-learning architectures enabled to operate on compressed samples of the image
2. Design of circuit architectures that allow the concurrence of compressive sensing and acceleration of on-chip inference
3. Design of a prototype chip demonstrating these principles
4. Integration of this prototype in a system for face/object recognition

^{*} This work is funded by the Spanish MICINN and the European Social Fund (ESF/FSE) through grant PRE2019-088478. This work has also been supported by European Regional Development Fund (ERDF/FEDER) through project RTI2018-097088-B-C31.

4 Summary of the Progress Made So Far

We have explored different machine-learning algorithms, and different architectures have been simulated in Matlab. Finally, we implemented the ‘one-vs.-all’ scheme of multiclass SVMs for multiclass classification. Three different facial datasets have been explored using the compressed learning approach. Different architectures have been tested for accurate and hardware efficient representation of compressed samples. Initially, a random binary measurement matrix was used to generate compressed samples and a very high recognition accuracy was observed. But this architecture suffers from a very large ADC resolution requirement. Therefore, an alternate architecture has been developed that requires a very small ADC (sigma-delta) resolution, but at the cost of reducing the accuracy. This new architecture uses a random ternary measurement matrix. The generation of random ternary measurement matrix can be done on the fly using elementary cellular automata. To enable a fair comparison with the state-of-the-art works, a figure-of-merit has been developed.

5 Current Status and Future Plan

Given that we have identified a hardware-efficient architecture for facial recognition, different circuit topologies are being explored for efficient hardware implementation for an on-chip sensor-processor system. The design and implementation of different analog/digital blocks in 110nm CMOS technology is to be carried out, achieving low power, small memory footprints, and area-efficient requirements.

The discussions with the mentor were very fruitful. These discussions established another basis for comparison of results of my measurement matrix with the standard Haar-transform matrix. Further elaborations resulted in the development of an alternate architecture based on Haar-transform matrix as the measurement matrix. The initial simulation results are very promising. In the near future, on-chip hardware implementation of the Haar-transform based measurement matrix is to be explored.

References

1. Calderbank, R., Jafarpour, S., Schapire, R.E.: Compressed learning: Universal sparse dimensionality reduction and learning in the measurement domain. Available at <http://dsp.rice.edu/files/cs> (2009)
2. Candès, E.J.: Compressive sampling. Proceedings of the International Congress of Mathematicians pp. 1433–1452 (2006)
3. Donoho, D.L.: Compressed sensing. IEEE Transactions on Information Theory **52**(4), 1289–1306 (2006)
4. Zarar, S.: Design of Energy-efficient Sensing Systems with Direct Computations on Compressively-sensed Data. Ph.D. thesis (September 2013)

First /Last name: **Yahjeb Bouha khatraty**

Title of the thesis: **Estimation of the quality of rice fields via intelligent systems**

Location of the thesis:

- ^a**Artificial Intelligence and Data Semantics Laboratory (LIASD), University Paris 8, Paris, France**
- ^b**Scientific Computing and Data Science Research Unit (CSIDS), University of Nouakchott Al-Aasriya, Nouakchott, Mauritania**

Expected start/end dates of the PhD: **2020/2023**

Names of the PhD co-supervisors : ^a**Nedra Mellouli** , ^b**Mamadou Tourad Diallo** and ^b**Mohamedade Farouk Nanne**

1 The context and objectives of the PhD

The application context of our project is climate change in Africa. It focuses on the management of water, an essential source of life, by applying Artificial Intelligence techniques. While water scarcity is increasing, fresh and clean water represents only 0.003% of all available water in the world. 70% of this water is used in agriculture and rice is one of the most water consuming crops in the world. However, at the same time, rice is a food staple for half the world's population. Thus, the increase in the world's population underlines the need to solve the water problem, but at the same time, we need to maintain sufficient production and good quality of rice yield. Our mission is to develop an intelligent system that combines the Internet of Things (IOT), a modern physical solution at the height of new technologies, and Artificial Intelligence, a modern and progressive application solution. The combination of these two solutions will result in an intelligent, scalable and resilient monitoring model to monitor, control and recommend water consumption related to irrigation by integrating other factors from the IOT such as temperature, humidity, evapotranspiration, vegetation indices from satellite images available on the agricultural region of Mauritania.

2 The research problem and its scientific basis

Mauritania has embarked on a policy of encouraging farmers to produce local rice by exploiting the country's local resources. Two major problems that stand in the way of the development of this activity are climate change, which is manifested by long periods of drought, and water scarcity. On the other hand, digital transformation is gaining ground in several areas and for the past three years, in favor of the field of digital and intelligent agriculture. Inspired by the very scarce state of the art on this topic, our problematic aims to develop more than a model, it is an intelligent monitoring system whose main tasks are:

- Monitor and control the quantity of water used in the rice fields.
- Monitor the state of the atmosphere (temperature, humidity, wind, rain, solar radiation), the state of the soil (temperature, humidity, nitrogen, phosphate, potassium (NPK) and pH) and the state of the plant (leaf moisture).
- Follow all the phenological phases of a plant and indicate if there are parasites infecting the plants (signal if there are exceptional cases that require an intervention in real time).
- Know the needs of the plants to have a good yield (find the relationship between the yield and the external factors).
- Predict the yield of an agricultural field and classify the different types of rice.

3 Tasks already completed

- Architectural conception of the rice field monitoring system.
- Data collection such as Normalized Difference Vegetation Index (NDVI), Normalized Difference Water Index (NDWI), Normalised Difference Snow Index (NDSI), Normalized Difference Moisture Index (NDMI), rainfall, Temperature, moisture, wind of Rosso Mauritania.

- Data processing such as statistical analysis of the data Analysis by decomposition (Trend, Seasonality and Noise) Analysis by clustering (partitioning and density-based methods).
- Prediction of data with machine learning models (ARIMA, SARIMAX, VAR, VARMAX and KNN)
- Prediction of data with Deep Learning models (LSTM and GRU).
- Publications: Khatraty, Y. B., Nauwynck, N. M., Diallo, M. T., & Nanne, M. F. REPORT ON ESTIMATING RICE FIELD QUALITY USING ARTIFICIAL INTELLIGENCE. 6th International Conference on Big Data Analytics, Data Mining and Computational Intelligence 2021 : 166-170.

Khatraty, Y. B., Nauwynck, N. M., Diallo, M. T., & Nanne, M. F. Prediction of satellite data using ARIMA and SARIMAX: the case of NDVI of Rosso Mauritania rice fields. The 4th International Electronic Conference on Remote Sensing. 2022

4 Work in Progress and Perspectives

- Generate synthetic time series data using GAN.
- Prepare a separate deep clustering model that uses GAN for feature extraction from time series data and prepares a new version of MDST-DBSCAN that uses adaptive density when searching for point neighbors and performs rollbacks for clustering error correction. The output of GAN will be the input to MDST-DBSCAN. Also, use dynamic time warping DTW instead of the distance used by classical DBSCAN.
- Optimize the best model found by several techniques (use of synthetic data(GAN), transfer learning, modification of hyperparameters...).
- Implementation of the sensor layer, network layer and application layer of the rice field monitoring system.
- Use data collected by the monitoring system and satellite data to predict rice yield for future years.
- Work on the services provided by this system such as monitoring and control of water quantity, alerting of exceptional cases requiring real time intervention and classification of different types of rice.

5 Results of the exchange with the mentor:

Since we lack data to train our prediction model, the monitor suggested using synthetic data and in particular the data generated by GAN. after some research, we found recent works that deal with the problem of data augmentation and imputation of missing time series data [1]–[4]. In the future, we will use one of the models found to augment our data.

On the other hand, since GAN is not yet well exploited in the field of clustering[5], we proposed to use it as a data feature extractor and its output will be used as input to an innovative version of the MDST-DBSCAN clustering algorithm.

Bibliography

- [1] X. Li, J. Luo, and R. Younes, "ActivityGAN: generative adversarial networks for data augmentation in sensor-based human activity recognition," in *Adjunct Proceedings of the 2020 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2020 ACM International Symposium on Wearable Computers*, Virtual Event Mexico, Sep. 2020, pp. 249–254. doi: 10.1145/3410530.3414367.
- [2] Y. Luo, Y. Zhang, X. Cai, and X. Yuan, "E²GAN: End-to-End Generative Adversarial Network for Multivariate Time Series Imputation," in *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence*, Macao, China, Aug. 2019, pp. 3094–3100. doi: 10.24963/ijcai.2019/429.
- [3] Y. Luo, X. Cai, Y. Zhang, and J. Xu, "Multivariate Time Series Imputation with Generative Adversarial Networks," p. 12.
- [4] M. Gupta and R. Beheshti, "Time-series Imputation and Prediction with Bi-Directional Generative Adversarial Networks," p. 8.
- [5] A. Alqahtani, M. Ali, X. Xie, and M. W. Jones, "Deep Time-Series Clustering: A Review," *Electronics*, vol. 10, no. 23, p. 3001, Dec. 2021, doi: 10.3390/electronics10233001.

Método para el descubrimiento de tópicos con aprendizaje profundo

Lezama Sánchez Ana Laura
BUAP / México

Mireya Tovar Vidal
BUAP / México

José Alejandro Reyes Ortiz
UAM / México

Abstract

In this report, we present the objectives of the thesis project as well as a presentation of the context and objectives of the PhD. On the other hand, the proposed methodology for the development of the project is presented, as well as the work carried out so far, the publications and future work, including the mentoring received in this doctoral consortium.

1 Objectives

General Objective The thesis has as main objective the proposal of a topic discovery method through deep learning and static and contextual word embeddings for the the following specific objectives:

1. Implement a computational module for the pre-processing of the texts used for topic discovery;
2. Develop a method for the automatic extraction of semantic relations of the type synonymy, hyponymy and hyperonymy in the collection of documents;
3. Apply and implement a model for embedding words with previously identified semantic relationships;
4. Propose a topic discovery method based on deep learning and the representation of texts;
5. Evaluate the topic discovery method in order to compare the results obtained with the existing works in the literature.

Exploitation Objectives

- Train researchers with a critical and interdisciplinary approach who are capable of conducting basic and applied research to produce new knowledge on the problems associated with language and knowledge engineering.

- Develop skills that allow them to create new paradigms that meet specific demands that society demands.
- Train high-level professionals capable of inserting themselves efficiently in the productive, academic or research sector where they can apply the skills and knowledge acquired.

A graduate of the doctorate in Language and Knowledge Engineering will be able to enter the workplace in educational institutions, research and industry. His areas of professional competence are those related to the treatment of information of any kind (image, voice, text, etc.). The graduate is capable of generating models that represent real problems related to the language and knowledge of human beings, in order to achieve a solution for them.

2 Summary of the Progress

The general methodology of this project thesis is composed of the following phases:

2.1 Corpora Gathering

The data that have been used within the literature, for example, the corpus in English from wikipedia (general corpus) for the embedding model on of words and for the discovery of topics the corpora *Reuters*¹ and *20 NewsGroups*².

2.2 Text pre-processing

This phase aims to clean the texts and normalize them to bring them to the same format. That is, part of a set of documents in natural language, its function will be to standardize the texts. The tasks of this phase for the corpora for topic discovery are:

¹<https://trec.nist.gov/data/reuters/reuters.html>

²<http://qwone.com/~jason/20Newsgroups/>

1. Remove non-ASCII symbols.
2. Convert uppercase to lowercase: in order to normalize the vocabulary.
3. Elements removal: XML, stopwords.
4. Part-of-speech (PoS) tagging: each item will be tagged with its grammatical category and morphological information.
5. Recognition of named entities: the parts of the text will be identified with semantic information such as names of people, places or time.

The tasks of this phase for the extraction of semantic relations on the general corpus are:

1. Content extraction in `<text>< /text>`
2. Remove non-ASCII symbols.
3. Segmentation

2.3 Semantic Relationship Extraction

This phase aims to identify the semantic relationships of synonymy, hyponymy and hyperonymy present in the general corpus in order to enrich the proposed embedding model. In this phase, the approach used to extract semantic relationships is through lexical-syntactic patterns.

2.4 Static Embeddings

this stage aims to obtain a word embedding representation model, based on semantic relations of the synonymy, hyponymy and hyperonymy type that will be the input to this stage, and as output a representation matrix will be obtained, following the methodology proposed in (Saedi et al., 2018) that developed a word embedding model extracting the semantic relations of *WordNet*.

2.5 Text Classification

The goal of this stage is to obtain the category for each text by training a convolutional neural network (CNN). Dividing the input corpus into two subsets, one for training and one for testing.

Evaluation this phase uses the test data set to evaluate the result of the text classification. For this evaluation we will use the metrics precision, completeness, F-measure and accuracy.

2.6 Topic Discovery

The goal of this phase is to discover the topics within the categories identified in the previous step through Latent Dirichlet Analysis (*LDA*) for topic discovery.

Evaluation in this phase, a comparison of the proposed method against the results of the state of the art will be carried out using the topic coherence metric.

3 Experimental Setup

Experiments were performed adding hyponymy, hyperonymy, synonymy, holonymy, and meronymy relationships from *WordNet* (Fellbaum, 2010) to the corpora *20 NewsGroups* and *Reuters*. The objective was to have a clear vision of the semantic value that each relationship provides. The conclusions reached were that hyponymy, hyperonymy, and synonymy relationships are those that provide the greatest semantic content and therefore, these three types of semantic relationships were searched within a corpus of English Wikipedia. As a result of these experiments, three articles were published (Sánchez et al., 2021a,c,b).

At present, word embeddings have been created with relationships extracted from Wikipedia. The generated model so far only has 20,000 relationships. For the evaluation of the word embedding model it has been used a convolutional neural network (CNN) and *20 NewsGroups* or *Reuters* like input corpus with a word embedding model created. Also, with the same corpus and the word embedding model (e.g., *GloVe*, *FastText*), we utilized the same procedure. The objective is to evaluate the quality grade of the proposed model. The classes obtained were evaluated with the metrics of accuracy, precision, recall, and F-measure. Yet, the results obtained show the need for a more knowledgeable embedding model.

We explored this technique for topic discovery has been employ with tweets in Spanish about COVID-19 and the obtained results will be exposed in the Future Technologies Conference congress 2022.

4 Future Research Plan

4.1 Contextual Embeddings

Static embeddings (e.g., *GloVe*, *FastText*) are context-independent: that is, there is only one (numerical) vector representation for each word. The

different meanings of the word (if any) are combined into a single vector. However, the Bidirectional Encoder Representations from Transformers (BERT) (Devlin et al., 2019) model creates embeddings that allow the word representation of multiple (multiple) vectors for the same word depending on the context in which the word is used. Therefore, BERT embeddings is context dependent. We decided to use BERT not only because it is easy to fine-tune, but it has also proved to be one of the most performing technologies in multiple natural language processing (NLP) tasks (Devlin et al., 2019; Lample and Conneau, 2019; Radford et al., 2018). This type of embeddings could enrich the semantic properties of these vector representations allow the meaning of texts to be encoded in such a way that similar texts are close in vector space for topic discovery. Thus, we decided to experiment with recent BERT-based models, such as BERTopic (Grootendorst, 2022). We also consider that enriching the input of such models with the semantic relationships of synonymy, hyponymy and hyperonymy present in the corpora could also envision an increase in performance and topic representability. The results obtained will be evaluated with the topic coherence metric and with an external corpus of three million documents to evaluate how coherent are the topics that are discovered with the proposed method.

4.2 Dynamic Topic Discovery

Traditional topic modeling techniques (e.g., LDA) are static in nature and do not allow sequential order documents to be modeled. Dynamic topic modelling (DTM) which was first introduced by Blei and Lafferty (2006) as an extension of the LDA, overcome this by modeling how topics may have evolved the time and scope of the performance reflects it. In this type of models, the temporal nature of topics does not influence the creation of global topics, since the same topic might appear across different times, albeit possibly represented differently. For this, we consider experimenting also with BERTopic (Grootendorst, 2022).

4.3 Topic Labelling

Topics are usually represented by a ranked list of words but this can be difficult and time-consuming for humans to interpret. Generating labels allows users to compare topical trends without having to align topics and to explore news collections by users who might not have the necessary linguistic

skills to do otherwise. Recent works (Allahyari and Kochut, 2015; Zosa et al., 2022) proposed ontological mapping-based methods that map topics to concepts in a language-agnostic news ontology. These concepts have labels in multiple languages that we use as topic labels. The ontology mapping can be approached as a multilabel classification task where a topic can be classified as belonging to multiple concepts.

We can use the IPTC Subject Codes as our news ontology³. This is a language-agnostic ontology designed to organise news content. This ontology has three levels with 17 high-level concepts, 166 mid-level concepts and 1,221 fine-grained concepts. Mid-level concepts have exactly one parent and multiple children. While Zosa et al. (2022) manually annotated a Finnish dataset, we can create our groundtruth for topic labelling with an semi-automatic matching between the concepts and the tokens of the topics, that could be checked by native speakers afterwards.

Evaluation We could use BERTScore (Zhang et al., 2019) to evaluate the labels generated by our models. BERTScore finds optimal correspondences between groundtruth topic labels and generated topic labels and from these correspondences, recall, precision, and F-score are computed. For each topic, we can compute the pairwise BERTScores between the groundtruth topic labels and the labels generated by the models and take the maximum score. We then average the scores for all topics report this as the model score.

5 Future Publications

The next publication that is planned to be carried out is in New Machine Learning and Deep Learning Techniques in Natural Language Processing where it will be presented from the extraction of relationships to the results obtained by taking as input the proposed embedding model and a corpus for discovery of topics to have tangible results on the performance of the proposed embedding model.

References

Mehdi Allahyari and Krys Kochut. 2015. Automatic topic labeling using ontology-based topic models. In *2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA)*, pages 259–264. IEEE.

³<https://iptc.org/standards/subject-codes/>

- David M. Blei and John D. Lafferty. 2006. [Dynamic topic models](#). In *Proceedings of the 23rd International Conference on Machine Learning, ICML '06*, page 113–120, New York, NY, USA. Association for Computing Machinery.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics*, pages 4171–4186.
- Christiane Fellbaum. 2010. Wordnet. In *Theory and applications of ontology: computer applications*, pages 231–243. Springer.
- Maarten Grootendorst. 2022. Bertopic: Neural topic modeling with a class-based tf-idf procedure. *arXiv preprint arXiv:2203.05794*.
- Guillaume Lample and Alexis Conneau. 2019. Cross-lingual language model pretraining. *arXiv preprint arXiv:1901.07291*.
- Alec Radford, Karthik Narasimhan, Tim Salimans, and Ilya Sutskever. 2018. Improving language understanding by generative pre-training.
- Chakaveh Saedi, António Branco, João Rodrigues, and Joao Silva. 2018. Wordnet embeddings. In *Proceedings of the third workshop on representation learning for NLP*, pages 122–131.
- Ana Laura Lezama Sánchez, Mireya Tovar Vidal, and José Alejandro Reyes-Ortiz. 2021a. A behavior analysis of the impact of semantic relationships on topic discovery.
- Ana Laura Lezama Sánchez, Mireya Tovar Vidal, and José Alejandro Reyes-Ortiz. 2021b. Enfoque para el descubrimiento de tópicos a través de wordnet. In *Coloquio de Investigación Multidisciplinaria 2021*.
- Ana Laura Lezama Sánchez, Mireya Tovar Vidal, and José Alejandro Reyes-Ortiz. 2021c. [Hypernyms-based topic discovery using LDA](#). In *Advances in Soft Computing - 20th Mexican International Conference on Artificial Intelligence, MICAI 2021, Mexico City, Mexico, October 25-30, 2021, Proceedings, Part II*, volume 13068 of *Lecture Notes in Computer Science*, pages 70–80. Springer.
- Tianyi Zhang, Varsha Kishore, Felix Wu, Kilian Q Weinberger, and Yoav Artzi. 2019. Bertscore: Evaluating text generation with bert. *arXiv preprint arXiv:1904.09675*.
- Elaine Zosa, Lidia Pivovarova, Michele Boggia, and Sardana Ivanova. 2022. Multilingual topic labelling of news topics using ontological mapping. In *European Conference on Information Retrieval*, pages 248–256. Springer.

Content-based image retrieval applied to radiological image databases

Thibault Sauron

LIPADE (EA 2517), Université de Paris (France) & Philips Healthcare (France)

Abstract. Our objective in this PhD project is to build a content-based image retrieval (CBIR) framework to enable queries by example in databases of medical images. Our work falls within the context of radiology and MRI imaging. Although CBIR and deep learning methods are performing increasingly better, their application in medical fields is challenging for various reasons. One such setback is the lack of medical databases annotated for CBIR. To solve this, an existing clinical database suitable for our task will be anonymized and reorganized to train neural networks with. Another challenge is to consider the fact that MRIs are multi-parametric images. As such, a query can't be a single image of a patient but must be composed of several images complementary of each other's. This leads to information fusion issues which need to be addressed as early as possible in the development of the method. Finally, the framework will be implemented in radiological databases to be used as decision support by radiologists in clinical routines.

Keywords: CBIR · Information Fusion · PACS · Contrastive Learning

1 Preamble

This thesis is funded by a scholarship from Région Île de France. It is supervised by both an academic laboratory and a company:

- Florence Cloppet (thesis director): Associate Professor (HDR), LIPADE (EA 2517) - SIP Lab, Université de Paris (France);
- Camille Kurtz: Associate Professor, LIPADE (EA 2517) - SIP Lab, Université de Paris (France);
- Antoine Olivier: Research Scientist, Philips Research France;
- Laure Fournier: Professor of radiology, Hôpital Européen Georges Pompidou (France).

The thesis has started on february 1st 2021 and will end on february 1st 2024. It has been going on for one year as the time of writing this document.

2 Introduction

Picture archiving and communication systems (PACS) are the database systems used in hospitals to store, edit and retrieve patient information (image acquisition, analysis and annotation, diagnosis editing, etc.). Although the amount of relevant information stored is significant, the queries that can be performed on these systems are limited and mainly restricted to keywords search on patient fields (patient name, date of examination, ...). The goal of this PhD project is to enable more fine-grained search on the PACS by the way of CBIR, using a patient image to retrieve similar ones, or patients containing similar lesions.

CBIR works by extracting meaningful features from images, in order to be able to compare them efficiently and find the closest one from a database. In our case we are focusing on visual similarity of patient's lesion. Means

of extracting these features can vary but state of the art is achieved by using latent representation of neural networks trained on a relevant task. Since domain shift from natural images to medical images is usually too important, neural systems, such as convolutional neural networks, need to be at least fine-tuned on medical images to be relevant in this medical context.

The two major contributions of this PhD project are (1) to develop a CBIR framework that is designed for medical images and all their specificities, as well as (2) to exploit the inherent multi parametric character of this mass of data.

3 Methodology

Although deep learning methods have seen real advances in recent years, their application on medical images raises several problems:

- **domain shift** from natural images is too high to enable plug and play transfer learning approaches from natural images;
- **annotated databases** are lacking in quantity and quality due to the complexity of gathering and annotating medical images;
- **validation** is extremely complex as results are hard to interpret, especially with retrieval evaluation which requires a reference ranking.

Therefore, developing deep learnings algorithms for medical images is not straightforward and raises several issues. One is the need of a significant amount of images, annotations relevant to the end task and the ability to assess the results. Considering this, our work focuses on three main points:

- anonymizing, curating and further annotating a database from a medical study[8] annotated with vi-

sual descriptors of the lesion, to use it as the backbone for training and testing CBIR approaches;

- developing methods to exploit the intrinsic multi-parametric character of medical images;
- developing a framework to train neural networks as well as CBIR systems to index and query medical image databases.

3.1 Databases

As mentioned previously, medical image databases are challenging to create. Even more so in the context of CBIR as the annotation needed are a reference ranking that can be compared to the result of the query. The EURAD[8] database addresses these problems as annotation comprises disease diagnostic and malignancy score but also the presence of 14 visual markers of the lesion. These markers can be used to have more fine-grained relevance score as two lesions of the same disease can look quite different. The visual marker annotations also correspond to what a radiologist would search for and use to make a diagnostic, as the database was created to assess a risk stratification decision rule using these markers. But there are several problems which make this database unusable as it is currently:

- it has not been properly anonymized, which legally prevents us from using it;
- it is a clinical database and was not designed to be used for machine learning, directory structures are not consistent, neither are MRIs sequences names, image quality, etc.
- it is not a public database and even though one goal is to make it accessible, because of all the previous reasons and others, it is not yet possible.

For all these reasons we are also working on another database: MICCAI-BRATS[6, 2, 1]. This database is a brain tumor segmentation challenge, with 1200 patients and 4 MRIs volumes per patient. Although it was not designed specifically for CBIR, segmentation masks can be derived into a proxy relevance score and it's being a public database allows verifiable evaluation and comparison to other approaches.

3.2 Information fusion

MRIs are multi-parametric images and are composed of several sequences : T1, T1Gd, T1CE, T2 to name a few. All sequence types allow determination of different tissue types. Even if not all of them are required at all time, one MRI exam needs several sequences acquired to make a diagnosis. Therefore, in order to build an efficient CBIR system, indexing only one sequence type is not sufficient.

One method of information fusion that has gained attention recently is contrastive learning. The basis of contrastive learning is to reduce the distance between the latent representation of an anchor positive samples while increasing it for negative samples. Positive samples can be

different views of the same subject[7], different transformations of the same image[3], an image and it's associated text[9, 5]... With contrastive training the multi-modality can be taken into account directly in the training phase.

We will investigate contrastive learning using patient information as a basis to determine positive couples. Unfortunately, from our preliminary experiments, the results so far are underwhelming, two sample close in the embedded space are not visually close and the database we use (BRATS) prevents us from having results that we can analyze. Adding image reconstruction loss to the neural network to emphasize the visual component of the neural network[4] and other methods to improve the results of the neural networks are considered. But building our own database has become a priority now, as it is what keeps us from moving forward the most.

3.3 CBIR on PACS

Proofs of concept on small databases with isolated and determined parameters will be made to show the capacity and usability of CBIR in medical contexts. However, the ultimate goal is to build a framework to be able to perform CBIR in the most generic way. The idea is to have a Neural Network backbone trained with our database generic enough to be able to extract meaningful features from most medical images and use it as a basis to index medical image databases. Since the goal is to find the most similar lesions, the Neural Network must be able to characterize not the whole image but only the lesions. This is a challenge, but a relatively simple solution would be to have the database annotated with ROIs so that the images can be cropped as desired to extract features only on the meaningful part. Also, depending on the performance of the Neural Network on specific databases, a fine-tuning step could be considered to adapt the neural network to the new images and lesions. Although contrastive learning can be made unsupervised, it would still require some amount of external information. How much information and of what type will need to be considered to achieve the best ratio between performance, genericity and time required to adapt database for the framework, especially as annotating a consequent number of patients is time consuming and requires specific knowledge.

4 Conclusion

This document introduced the ultimate goal of this PhD thesis and defined as well several intermediate steps that are required in order to achieve it. These steps include: the building of a database suited for CBIR and the training method of a Neural Network on multi-parametric images in a semi-supervised way, to deal with our information fusion issues.

In the next two years these specific milestones will need to be completed. We then aim to work towards solving real cases scenarios, in clinical routines.

References

- [1] Ujjwal Baid et al. “The RSNA-ASNR-MICCAI BraTS 2021 Benchmark on Brain Tumor Segmentation and Radiogenomic Classification”. In: (2021). arXiv: 2107.02314v1.
- [2] Menze BH et al. “The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)”. In: *IEEE transactions on medical imaging* 10 (2015). URL: <https://pubmed.ncbi.nlm.nih.gov/25494501/>.
- [3] Ting Chen et al. “A Simple Framework for Contrastive Learning of Visual Representations”. In: PMLR, 2020. URL: <http://proceedings.mlr.press/v119/chen20j.html>.
- [4] Kazuma Kobayashi et al. “Decomposing normal and abnormal features of medical images for content-based image retrieval of glioma imaging”. In: *Medical Image Anal.* (2021). URL: <https://doi.org/10.1016/j.media.2021.102227>.
- [5] Alec Radford et al. “Learning Transferable Visual Models From Natural Language Supervision”. In: PMLR, 2021.
- [6] Bakas S et al. “Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features”. In: *Scientific data* (2017). URL: <https://pubmed.ncbi.nlm.nih.gov/28872634/>.
- [7] Matthew Schultz and Thorsten Joachims. “Learning a Distance Metric from Relative Comparisons”. In: MIT Press, 2003. URL: <https://proceedings.neurips.cc/paper/2003/hash/d3b1fb02964aa64e257f9f26a31f72cf-Abstract.html>.
- [8] Isabelle Thomassin-Naggara et al. “Ovarian-Adnexal Reporting Data System Magnetic Resonance Imaging (O-RADS MRI) score for risk stratification of sonographically indeterminate adnexal masses”. In: *JAMA network open* 1 (2020).
- [9] Yuhao Zhang et al. “Contrastive Learning of Medical Visual Representations from Paired Images and Text”. In: *CoRR* (2020). arXiv: 2010.00747. URL: <https://arxiv.org/abs/2010.00747>.

Automatic and model-free learning of semantic-structural links of fields in a corporate documents

Ibrahim Souleiman^{+,*}, Mickaël Coustaty⁺
Aurélie Joseph^{*}, Vincent Poulain d'Andecy^{*}, Jean-Marc Ogier⁺

^{*} Yooz

1 Rue Fleming, 17000 La Rochelle, France

Email: {ibrahim.souleimanmahamoud,aurelie.joseph,vincent.poulaindandecy}@getyooz.com

⁺ La Rochelle Université, L3i

Avenue Michel Crépeau, 17042 La Rochelle, France

Email: {mickael.coustaty,jean-marc.ogier}@univ-lr.fr

Abstract—This project is a continuation of the work initiated in the last three years between the company YOOZ and the L3i and in the joint laboratory ANR IDEAS. The general objective is to propose a generic system of information extraction from document streams to automatically process a heterogeneous set of documents from companies. The specific objective of this project is to use state-of-art method to automatically detect the concepts of a document. This thesis, thus aims at applying a methodology and a theoretical model allowing the analysis of documents of any kind to extract precise information necessary for all levels of document processing.

I. INTRODUCTION

In the age of digitalization, the automatic processing of administrative documents has become both an economic and technological challenge. Automatic Document Reading (ADR) is a software solution that automatically reads these scanned documents and extracts the useful information to inform the information systems and process them quickly. For the past 30 years, ITESOFT and now YOOZ has been developing information extraction and classification tools that seek to reproduce the capabilities of humans while being more efficient. However, while the documents of choice for a company were invoices, forms, financial regulations, and mail, the diversity has extended to other types of documents that are much more varied but also more ambiguous (e.g. purchase orders, estimates, deliveries, etc.). Even within the documents of choice, variability is increased by the increase and internationalization of B2B exchanges. Document analysis approaches by templates or keywords, adapted to small volumes, are no longer sufficient. In terms of new technologies, Deep Learning or incremental learning systems have become essential but do not offer reliable industrial solutions to the problems of real datasets (strong imbalance between learning classes, lack of classes and models can online learning, etc.).

II. OBJECTIF AND CHALLENGES

The objective of this thesis can be broken down into 4 axes:

- Learn and reconstruct automatically the semanticstructural link of fields in a document. This should allow the development of a general model applicable to documents with similar semantics
- Extract data without a priori manual model (business rules entered by an expert for each client) on a large and multilingual vocabulary. The models will be learned from examples of results
- Propose methods with little or no parameterization
- Satisfy a low processing time and respect the industrial constraints on error minimization

The two main technological challenges are :

- Simplicity and automation of learning that can be achieved from naive examples, within reach of an end user.
- Continuous adaptation of classifiers as data is discovered.

III. RESEARCH

In this section we will describe the models we want to implement to bring solutions to the problem defined before. Document classification and information extraction for a human requires some expertise, often related to the business or domain. Document sorting often relies on the use of taxonomies (document classes) and on similarities with previously encountered documents, in order to decide which class to assign. From a machine learning point of view, this relies on the recognition of certain patterns in the documents, the position of these patterns and the presence of certain relevant words (keywords) that allow to match and classify a set of documents or a set of words. Many solutions have been proposed in the last twenty years.

The extraction model will have input the prediction of the class because the information to extract are related by the class of document. The classification and extraction model may or may not use the same architecture for example they may use a combination of expert system, incremental learning and deep learning.

There are several ways to extract the relevant information inside a document (e.g. Key information extraction, Visual question information, etc). For our task when we extract information, it will be through questions and answers. The needs of our client are diverse, some clients want to extract information common to all customers (i.e. total amount, invoice number, etc) and others want to extract specific information (i.e. membership number).

Several works have focused on the textual part of the documents others use a multimodal neural network able to learn from a lexical folding of the text extracted by character recognition [1] and visual features of the image (MobileNetV2 [2]).

Despite the good performances of these deep learning techniques, they have their limits in an industrial context. State-of-the-art methods have 100 million of parameters, are time consuming and it's difficult to interpret the result (i.e. some client need to when more information about the why this output). Also the company which is confronted with huge document flows can be treated with other types of documents (e.g. new classes of documents, new structure of documents) Incremental learning brings a solution to this problem.

This project is at the interface between incremental learning in order to adapt the system to the evolution of the corpus, deep learning in order to obtain a most relevant and generic characterization for large volumes of documents, and finally the expert systems which make it possible to propose rules specific to the end user's business. To our knowledge, the joint study of these three typologies of approaches, in particular in the context of processing large flows of documents, has never been carried out. This thesis project therefore aims to establish a complete state of the art on these themes, define the approaches relevant to our problem and study their combination. Finally, to evaluate and validate all of this work, we will rely on a set of data sets and the classic metrics from the literature (Recall, Precision, F-Measure, ANLS, Exact Match). Several databases are freely available in order to be able to pretrain the models and compare them. For the best known and largest classification dataset, which currently serves as a worldwide reference, is the RVL-CDIP [3], DocVQA [4] and others. These and other datasets will allow us to compare ourselves to other state-of-the-art methods.

IV. STAKEHOLDERS IN THE PROJECT

Title of the thesis : Automatic and model-free learning of semantic-structural links of fields in a document

Student in thesis: Ibrahim Souleiman Mahamoud (70% of time in Laboratory L3i - 30% in YOOZ Company)

Director: Jean-Marc Ogier, L3i, University of La Rochelle

Scientific supervisor: Mickaël Coustaty, L3i, University of La Rochelle

Industrial supervisor : Aurélie Joseph, Yooz, Aimargues

Starting Date : 15/04/2021

Finalization date : 14/04/2024

V. CV

A. EDUCATION

- PhD Extraction information from documents(04/2021-04/2024) University of la Rochelle and YOOZ company
- Post-graduate Artificial intelligence (09/2019-09/2020) École de centrale Marseille
- Master's degree Artificial intelligence and machine learning (09/2018-09/2019) University Aix-Marseille
- Bachelor's degree Database and Software (04/2015-04/2016) University of Djibouti
- High-School Diploma (09/2012-04/2013) High-School Baballa in Djibouti City

B. PUBLICATIONS

- DAS-2022 — QALayout: Question Answering Layout based on multimodal Attention for visual question answering on corporate Document
- ICDAR-2021 — Multimodal Attention-based Learning for Imbalanced Corporate Documents Classification
- EGC-2021 (09/2019-09/2020) — Apprentissage multimodal basé sur des modèles d'attention pour la classification de documents dans un contexte déséquilibré

C. Work Experience

- Fixed-term contracts (09/2020-04/2021) Company YOOZ — Documents Classification
- Internship Company YOOZ — Documents Classification
- Internship in Laboratory TALEP (Aix-Marseille University) (03/2019-07/2019) —Multimodal comprehension on the GuessWhat?!
- Internship Laboratory Qarma (Aix-Marseille University) (06/2018-07/2018) — Artificial Intelligence for shufimi game

REFERENCES

- [1] N. Audebert, C. Herold, K. Slimani, and C. Vidal, "Multimodal deep networks for text and image-based document classification," in *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer, 2019, pp. 427–443.
- [2] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 4510–4520.
- [3] A. W. Harley, A. Ufkes, and K. G. Derpanis, "Evaluation of deep convolutional nets for document image classification and retrieval," in *2015 13th International Conference on Document Analysis and Recognition (ICDAR)*. IEEE, 2015, pp. 991–995.
- [4] M. Mathew, D. Karatzas, and C. V. Jawahar, "Docvqa: A dataset for vqa on document images," 2021.

ICPRAI 2022 Doctoral Consortium

Christoph Wehner^[0000–0003–0421–4113]

University of Bamberg, Kapuzinerstraße 16, 96047 Bamberg, Germany
`christoph.wehner@uni-bamberg.de`

1 Research Plan

The title of my thesis is "Explainable and Interactive Link Prediction on Knowledge Graphs". The thesis is supervised by Prof. Dr. Ute Schmid. I started the thesis at the 01.01.2022, and I hope to finish it until the 01.01.2025.

1.1 The Project KIProQua

Currently, I am working on the project KIProQua. I research novel AI-driven approaches to root-cause detection of defects in the production chain of electric powertrains. I aim to combine sub-symbolic machine learning methods and symbolic AI to incorporate interactively expert knowledge into root-cause detection methods.¹

1.2 Arising Challenges

Several exciting research challenges arise through this project. First, the use case of KIProQua provides me with multiple large-scale knowledge graphs of the production chain. Those knowledge graphs are notoriously incomplete, and the entities holding the root cause for the defects are not trivial to detect.

Furthermore, it is crucial to the quality compliance that the model explains why it believes the entities to be the root cause of the defect.

Thus, experts shall be able to incrementally give feedback on the models' prediction and explanation to improve the models' prediction while using the system in their daily work.

1.3 First Steps Taken of my Doctoral Degree Journey

I started my doctoral degree by learning about the use cases of the project KIProQua. Furthermore, I am conducting a literature review on explainable link prediction in knowledge graphs and how to make them interactive. I am working on a survey regarding that topic in cooperation with a college of mine at BMW. Finally, my first paper, titled "Explainable Online Lane Change Predictions on a Digital Twin with a Layer Normalized LSTM and Layer-wise Relevance Propagation", was published at the IEA AIE 2022. Its scientific contribution is extending LRP [2] to layer normalized LSTM's [1].

1.4 Research Idea

I aim to develop a novel method that allows for interactive learning of a link prediction model by leveraging the models' explanations.

Research in this direction is significant since, in domains where labelled data is sparse, knowledge graphs are incomplete, and labelling by domain experts is

¹ Feel free to visit our website for further information: <https://www.uni-bamberg.de/en/cogsys/research/projects/bayvfp-project-kiproqua/>

expensive, it can provide an accessible, controlled, and safe way of improving the link prediction models' performance.

Research on explainable link prediction is an emerging topic [4, 6, 7]. And using those explanations to improve the model was identified as a novel research frontier by the link prediction literature [5] in 2020. However, to the best of the authors' knowledge, no effort in this direction has been taken ever since. Thus, combining explainability and interactive learning in the link prediction domain will be an innovation.

I want to approach this topic by solving real-world problems arising in the project KIProQua, resulting in method and applied papers published in top-tier conferences.

1.5 Next steps

The fruitful discussions with my mentor Oriol Ramos Terrades heavily inspire my next steps. I aim to combine the method from [4] called *MINERVA* with the preference-based human-in-the-loop reinforcement learning (RL) approach from [3]. *MINERVA* models the link prediction task of the tail entity given a head entity and a missing relation as a pathfinding problem. Knowledge graph nodes are treated as states, and relations are the actions or paths to transition from one state to the other. An RL agent is trained to walk from the head entity the optimal path, given a missing relation, through the knowledge graph towards the correct tail entity [4]. The optimal path is the symbolic explanation of the models' behavior. By adding the preference-based human-in-the-loop utility function from [3] to the RL agent, interactivity is achieved. In the following steps, the RL agent shall be enabled to consider literals. Finally, a version with a rule-based human-in-the-loop utility function, instead of a preference-based utility function, shall be implemented. This way, I hope to get closer to my final goal of achieving a doctoral degree.

References

1. Ba, J.L., Kiros, J.R., Hinton, G.E.: Layer normalization (2016)
2. Bach, S., Binder, A., Montavon, G., Klauschen, F., Müller, K.R., Samek, W.: On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation. *PLoS ONE* **10**(7), e0130140 (07 2015). <https://doi.org/10.1371/journal.pone.0130140>
3. Christiano, P.F., Leike, J., Brown, T., Martic, M., Legg, S., Amodei, D.: Deep reinforcement learning from human preferences. vol. 30. Curran Associates, Inc. (2017)
4. Das, R., Dhuliawala, S., Zaheer, M., Vilnis, L., Durugkar, I., Krishnamurthy, A., Smola, A., McCallum, A.: Go for a walk and arrive at the answer: Reasoning over knowledge bases with reinforcement learning. *OpenReview.net* (2017)
5. Futia, G., Vetrò, A.: On the integration of knowledge graphs into deep learning models for a more comprehensible ai—three challenges for future research. *Information* **11** (2020). <https://doi.org/10.3390/info11020122>
6. Lecue, F.: On the role of knowledge graphs in explainable ai. *Semantic Web* **11**, 41–51 (2020). <https://doi.org/10.3233/SW-190374>
7. Ma, W., Zhang, M., Cao, Y., Jin, W., Wang, C., Liu, Y., Ma, S., Ren, X.: Jointly learning explainable rules for recommendation with knowledge graph. pp. 1210–1221. Association for Computing Machinery (2019). <https://doi.org/10.1145/3308558.3313607>

Weakly Supervised Learning on Histopathological Images for Immune Profiling

Qinghe Zeng

Laboratoire d'Informatique Paris Descartes (LIPADE, EA 2517)
Université Paris Cité
qinghe.zeng@etu.u-paris.fr

Title of Thesis

Intelligent Exploration of Whole Slide Image for Digital Histopathology

Location of Thesis

Paris, France 75006

Starting and Expected Finalization Date of the PhD

17/11/2019 - 17/11/2023

Supervisors of Thesis (Multidisciplinary PhD)

Nicolas Lomenie / Associate Professor, Eng, PhD, HDR / Université de Paris
Julien Calderaro / Professor, PhD, MCU-PH, MD / Hôpital Henri Mondor – APHP
Christophe Klein / IR INSERM / Centre de Recherche des Cordeliers UMRS 1138

1 Context and Objectives of the PhD

Cancer is one of the leading causes of death in modern society, with 18.1 million new cases and 9.5 million related deaths worldwide in 2018 [1]. Thankfully, cancer treatments are evolving, from radiotherapy, chemotherapy, targeted therapy, and more recently, checkpoint inhibitors (namely immunotherapy). However, to achieve a better prognosis, these treatments should be given in combination or individually to each patient depending on the cancer type/subtype, grading, as well as immune status especially for immunotherapy. Computational pathology is leading a revolution in cancer diagnosis, prognosis and treatment response prediction thanks to the promotion of artificial intelligence, especially deep learning, making personalized medicine possible in clinical practice.

Immunotherapy is one of the most promising cancer therapies, which fights efficiently against tumors by stimulating the immune system. However, only a subset of patients can benefit from immunotherapy due to the differentiation of immune profiles. Therefore, identifying these potential patients is a clinically essential issue. Several gene signatures have been shown to have significant predictive values in many cancers, including hepatocellular carcinoma (HCC) which is the most common primary liver cancer. However, gene profiling techniques are highly sensitive to sample quality and prone to standardization issues. Limited by their high cost and time-consuming nature, they may not be easily available, especially in underdeveloped regions. On the other hand, histological slides are readily available in clinical routine and contain information for neoplastic diagnosis and outcome prediction. Thanks to the invention of whole slide imaging, computational pathology offers an opportunity to quantify and standardize digital histological analysis for reduced subjectivity with increased efficiency and reproducibility. Therefore, we aimed to develop a clinical-grade workflow for immune profiling directly from histological slides, to aid immunotherapy response and prognosis prediction.

2 Summary of the Progress Made So Far

We reviewed the existing approaches for assessing tumor-related immune status, such as spatial pattern recognition and tumor-infiltrating lymphocyte scoring, as summarized in our paper published in the British Journal of Pharmacology [2]. However, most of these approaches necessitate substantial engineering at the cellular level and/or expensive immunochemistry or manual annotation, with the exception of a convolutional neural network-based approach, which needs only coarse tumor annotation but performs suboptimally for HCC gene signatures. In our work

published in the Journal of Hepatology [4], we tested 3 different deep learning approaches to predict the immune status (activation or suppression) of 6 gene signatures previously proved to be related to immunotherapy response and survival [5]. An attention-based Multiple Instance Learning (MIL) approach, namely CLAM [3], is shown to significantly outperform the others. In another series of external validation, using different gene profiling techniques and different staining protocols, our models showed excellent generalizability of 0.817, 0.810, 0.850, 0.823, 0.810, 0.921 for the 6 investigated signatures, respectively. Our proof-of-concept study suggests that deep learning has the ability to capture morphological features of immune predictive value.

3 Future Research Plan for the Last Part of the Thesis

A necessary next step will be to validate the applicability of our models trained on resection samples in a board biopsy series, as biopsy is the only sample type for advanced HCC patients. Unlike resection samples, biopsies contain only a small amount of tissues aspirated from the tumor which are usually damaged, and therefore may have more artifacts and different morphological appearance. Domain adaptation methods will be investigated for this objective. One perspective is the direct prediction of gene expression levels. In our previous study, we aimed to predict the activation of six immune-related gene signatures. However, these labels are highly dependent on unsupervised clustering methods, which are difficult to be validated. On the other hand, the prediction of gene expression levels is a very challenging task for which there are no satisfactory methods for the time being. By adding gene expression levels in the same signature as additional outputs, the models are expected to benefit from the “real” labels and capture better features in a multi-task learning mechanism. We will also investigate recent techniques of self-supervised learning and spatial information preserved techniques for this task. In addition, if we are able to collect clinical data, we will also validate the predictive value of our model on immunotherapy response and survival.

Another important perspective will be to study the tumor heterogeneity. Our current workflow is able to produce a patch-level attention map for interpretability. We plan to validate this attention map using spatial transcriptomics techniques, that is, gene expression profiling data of about 5000 small spots arranged evenly on a histological image. As a histological image is usually extremely large (100,000 x 100,000 px) to load into a consumer-grade GPU, a common preprocessing step to segment it into small images (ex. 256 x 256 px), so-called patches. To discover new biomarkers or morphological patterns, more finer-grained interpretability will be needed. We will investigate novel weakly supervised learning models, such as the nested MIL [6] to predict cell- or pixel-level gene expression directly from histological images with only image-level labels.

Eventually, we would like to deploy this workflow to a no-coding user-friendly platform, to boost the use by physicians, especially those without deep learning skills.

References

1. Cancer statistics. National Cancer Institute. (n.d.). Retrieved February 15, 2022, from <https://www.cancer.gov/about-cancer/understanding/statistics#:~:text=Cancer>
2. Klein, C., Zeng, Q., Arbaretaz, F., Devèvre, E., Calderaro, J., Lomenie, N., & Maiuri, M. C. (2021). Artificial intelligence for solid tumour diagnosis in digital pathology. *British Journal of Pharmacology*, 178(21), 4291-4315.
3. Lu, M. Y., Williamson, D. F., Chen, T. Y., Chen, R. J., Barbieri, M., & Mahmood, F. (2021). Data-efficient and weakly supervised computational pathology on whole-slide images. *Nature biomedical engineering*, 5(6), 555-570.
4. Zeng, Q., Klein, C., Caruso, S., Maille, P., Laleh, N. G., Sommacale, D., ... & Calderaro, J. (2022). Artificial intelligence predicts immune and inflammatory gene signatures directly from hepatocellular carcinoma histology. *Journal of Hepatology*.
5. Sangro, B., Melero, I., Wadhawan, S., Finn, R. S., Abou-Alfa, G. K., Cheng, A. L., ... & El-Khoueiry, A. (2020). Association of inflammatory biomarkers with clinical outcomes in nivolumab-treated patients with advanced hepatocellular carcinoma. *Journal of hepatology*, 73(6), 1460-1469.
6. Fuster, S., Eftestøl, T., & Engan, K. (2021). Nested multiple instance learning with attention mechanisms. *arXiv preprint arXiv:2111.00947*.