UMR IRISA

Activity Report 2021

Team INTUIDOC

Intuitive user interaction for documents

D6 – Media and Interactions

CentraleSupélec · cnrs · ens rennes · IMT Atlantique Bretagne-Pays de la Loire École Mines-Télécom · Inria · INSA RENNES · Université Bretagne Sud ubs · UNIVERSITÉ DE RENNES 1

1 Team composition

Researchers and faculty

Eric Anquetil, Professor, Insa, head of the team
Bertrand Coüasnon, Associate Professor, Insa, HDR
Nathalie Girard, Associate Professor, Univ. Rennes 1
Aurélie Lemaitre, Associate Professor, Univ. Rennes 2, HDR
Ivan Leplumey, Associate Professor, Insa
Yann Ricquebourg, Associate Professor, Insa

Associate members

Jean Camillerapp, Retired Professor
Yann Soullard, External Colloborator, Associate Professor, Univ. Rennes 2

Research engineers, technical staff

Simon Bouvier, Insa Research Engineer, until May 7th
Kwon-Young Thomas Choi, Insa Research Engineer, from April 19th to August 18th
Iwan Le Floch, Insa Research Engineer
Pauline Nerdeux, Insa Research Engineer

PhD students

Killian Barrère, Insa PhD student
Simon Corbillé, Univ. Rennes 1 PhD student
Martin Dornier, CIFRE/Insa PhD student
Camille Guerry, Insa PhD student
Florent Imbert, Insa PhD student
Clément, Insa PhD student
Arnaud Lods, CIFRE/Insa PhD Student
William Mocaër, Insa PhD student
Timothée Neitthoffer, CIFRE/Insa PhD Student, from May 1st
Solène Tarride, CIFRE/Insa PhD Student

PostDoc

Omar Krichen, Insa Post Doc
Wassim Swaileh, Insa Post Doc, from October 4th

Administrative assistant

Nadia Derouault

2 Overall objectives

2.1 Overview

The Intuidoc team focuses its work on handwriting, gesture (2D and 3D), and documents under various aspects: analysis, recognition, composition, interpretation. We are also interested in human-document interaction and graphical/gestural man-document interaction. This research relates to the handwriting and the documents under different forms: manuscript, printed paper form, pen-based and touch-based interaction, graph, images, heterogeneous documents, etc.

The roadmap of the IntuiDoc team is on the frontier of several research axes: Pattern recognition, Machine-Learning, Artificial Intelligence, Human-Machine Interaction, Uses and Digital Learning. The aim is to explore new scientific challenges of the domain of the Human-Document Interaction with a specific focus on interactive, incremental and evolving learning based on the integration of the user in all the processes of analysis and decision making.

Today, four major emerging scientific axes are investigated with strong partnerships with national and international laboratories and companies:

- "On-line" evolving cross-learning of 2D (touch and pen–based) and 3D gestures (Kinect and Leap Motion);
- "On-line" analysis of drawing, sketching and handwriting with pen-based tablet for digital learning (e-education);
- Interactive learning of document structure without ground-truth;
- Document collection analysis for big-data.

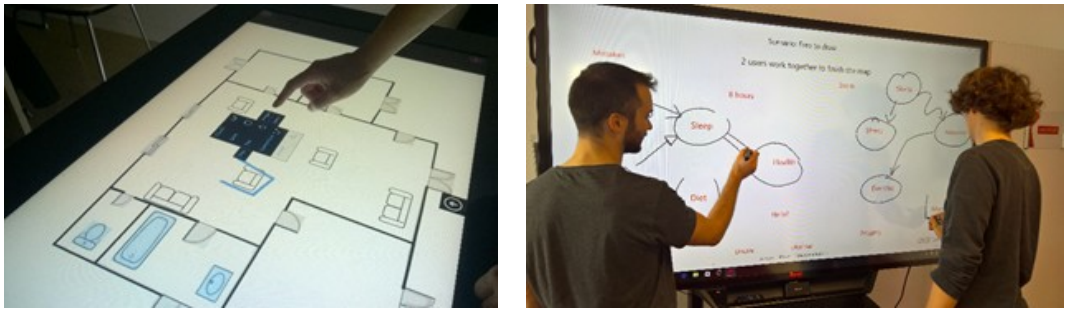
2.2 Scientific foundations

2.2.1 On-line evolving cross-learning of 2D and 3D gestures.

2D evolving recognizer for gesture commands With the increasing use of touch and pen-based sensitive screens, human-computer interactions are evolving. New interaction methods have been designed to take advantage of the new potential of interaction offered by these interfaces. Among them, a new concept has recently appeared: to associate commands to gestures. Those gesture commands enable users to execute various actions simply by drawing symbols. This new man-machine interaction can be used for on-line composition of complex documents such as electrical sketches or floor plan. In order to use such gesture commands, a recognition system is required. For users to easily memorize more than a dozen of gesture commands, it is important to enable gesture set customization. The classifier used to recognize drawn symbols must hence be customizable, able to learn from very few data, and evolving to learn new classes on-the-fly and improve during its use. The objective of this work is to obtain a gesture command system that cooperates as best as possible with the user, learning from its mistakes without soliciting the user too often. Gesture commands lead to a cross-learning

situation where the user has to learn and memorize the gestures, and the classifier has to learn and recognize drawn gestures. We study the impact of different strategies to supervise the online training of an evolving recognizer for gesture commands, and how to optimize this cooperation between the user and the recognition system. In particular, we design an inner confidence measure to solicit the user when some data samples don't fit the classifier model, and that it will be very gainful to learn from it.

Multi-touch gesture recognition Due to the recent prevalence of multi-touch devices, multi-touch gesture recognition has gained a large interest in the last decade. Unlike mono-touch gesture recognition which tracks the movement of a single point of input, multi-touch gesture often tracks many points of contact in parallel as they appear, move and disappear. The recognition for multi-touch gestures is challenging because of the complex chronological relation between the fingers' trajectories. We explore new methods for modelling the shape, relative temporal and motion information in multi-touch gesture by a model of graph and graph embedding approach. In our future work we aim at developing a strategy to detect the pattern of multi-touch gesture at runtime, to be able to address direct manipulation by command gesture.



(a) (b)

Figure 1: Multi-Touch and Multi-User Interaction

Multiple user freely-drawn sketch recognition and 3D action gesture recognition Another scientific challenge is also to address large multi-touch display that allows multiple users to simultaneously interact in the same context and work together. Indeed, many researches and commercial products propose tangible interfaces which support simultaneous participation of multiple users. This is a really new research topic to automatically recognise and interpret in real time the freely-drawn sketch of multiple users.

Finally, in this axe, we investigate the validity of transferring the expertise on hand-drawn symbol representation [DA13] to recognise 3D action gesture. This new research topic will be conducted in collaboration with MIMETIC project team of Inria. We base this proposition on the observation that patterns produced by a human motion, in particular 2D hand-drawn symbols and 3D actions, share several important properties. They are both governed by kinematic constraints that must be considered while

[DA13] A. Delaye, E. Anquetil, "HBF49 feature set: A first unified baseline for online symbol recognition", *Pattern Recognition 46*, 1, 2013, p. 117 – 130.

modelling such human motions. We hypothesise that both recognition problems could be addressed in similar ways.

2.2.2 Artificial Intelligence for e-education

This research axis is more recent in the team. It focuses on the design of artificial intelligence engines for e-education. It is a very active line of research since it is associated with 5 major research projects over the past 5 years with a funding budget of 1.456 Million euros.

2.2.3 Handwriting analysis for digital learning at school

The scientific problem we tackle here is to quantitatively evaluate a cursive handwriting with respect to a reference model and recommendations of a teacher. In order to be able to teach children how to write, we must be able to analyse their handwriting, to evaluate if the letters, words, sentences are correctly written, and to detail which aspects of the child handwriting do not correspond to the teacher models (corrective feedback). This problem is completely different from the classical task of character recognition, where the challenge is to determine to which class the data samples belongs.

Our objective is to be able to analyse, qualify and evaluate handwriting, with regards to reference models, and for multiple distinct aspects like: shape (for legibility), drawing direction and order (for ductus), speed and fluidity for instance. We use an analysis system based on an evolving fuzzy classifier. It allows to easily define reference models from few data samples to customise "on the fly" the writing exercises to the children. Then, the analysis system can be used to evaluate drawn gestures, regarding a specific feature set, and finally give a confidence score.

The ANR Joint laboratory (LabCom: Script&Labs) between IntuiDoc and Learn&Go The axe of Artificial Intelligence for e-education has been launched with the IntuiScript project founded by the French government as part of innovative national projects (BPI-PIA2). IntuiScript targets towards offering an advanced digital writing experience at school by using tablets and tactile digital devices (with finger touch and stylus). This project was structured around the conception of a digital workbook to help teachers and children from three to seven years old during the handwriting learning process:

- it allows children to work in autonomy with an on-line and real time feedback;
- it proposes automatically pedagogical exercises that are adapted to children difficulties based on the automatic analysis of children writing;
- it provides a precise off-line analysis of children writing (i.e. order, direction, shape) to help teachers to understand children writing skills and difficulties.

This project was based on a user-centred design approach that includes several cycles of conception followed by experiments. Therefore, feedback of children and teachers

related to these experiments have been used to improve the education scenario. More than 1,000 primary school students from Brittany have taken part to the experiments in the project.

This four years project was a real success. It resulted in the launch of the product "Kaligo" today distributed in schools by the company Learn & Go.

Figure 2: Handwriting analysis for digital learning

With the success of IntuiScript project, the IntuiDoc team and Learn&Go company created the "Script&Labs" LabCom to innovate on Digital Learning.

The scientific principles of the joint laboratory belong to artificial intelligence (pattern recognition, Machine Learning) and man-agent interaction. Interpretation, adaptation and learning are the heart of its researches, aiming to conceive automated interpretation systems for children productions : writing, arithmetic operations, geometric figures. This scientific know-how form the foundation of new solutions in Digital Learning. leading to more independence and customisation for each student learning process.

The results of the joint laboratory will directly be used in innovating educational modules focusing in active learning, handwritten input, analysis and personalised help via immediate feedback on student production [SAB17].

Three lines have been currently chosen :

- writing learning for young children
- learning numbers and first arithmetic calculations
- geometry learning by “generative drawing”

Beyond these three lines, LabCom Script&Labs helps working on “active and collaborative learning from secondary school to higher education” through e-Fran « Actif » et Cominlabs « e-Fil » projects.

The “ScriptAndLabs” joint laboratory is founded by ANR for four years (n° ANR-16-LVC2-0008-01).

In 2019, we won with Learn&Go company a new project from the Ministry of Education (DNE) as part of the Innovation - Artificial Intelligence Project (P2IA). The

[SAB17] D. Simonnet, E. Anquetil, M. Bouillon, “Multi-Criteria Handwriting Quality Analysis with Online Fuzzy Models”, *Pattern Recognition*, 2017, `https://hal.archives-ouvertes.fr/hal-01515397`.

objective is to design a software environment based on artificial intelligence for helping to learn French (writing / spelling) for teachers and students in cycle 2. In this project, we are responsible for designing and developing the automated analysis engine of handwriting to identify the spelling/graphemes mistakes of pupils in CP, CE1 and CE2.

"On-line" analysis of drawing for new learning strategies based on "generative drawing" One of the key topics in this axe is how to encourage new learning strategy based on "generative drawing" using pen-based numerical devices. The goal is to improve the learning of students by considering "learning" as a generative activity. In this scope, the potential induced by pen-based tablet is really interesting. The goal is to investigate how we can automatically generate intelligent "corrective" or "predictive" feedbacks to a user during his drawing process: summarizing, mapping, drawing, sketching... We base this work on the visual grammar CD-CMG [MA09] (Context Driven Multi-set Grammar), to model the domain knowledge and interpret the hand-drawn sketches on the fly. We adapted this grammar to the Geometry domain to cover the concepts taught in middle-school (cf. Figure 3).

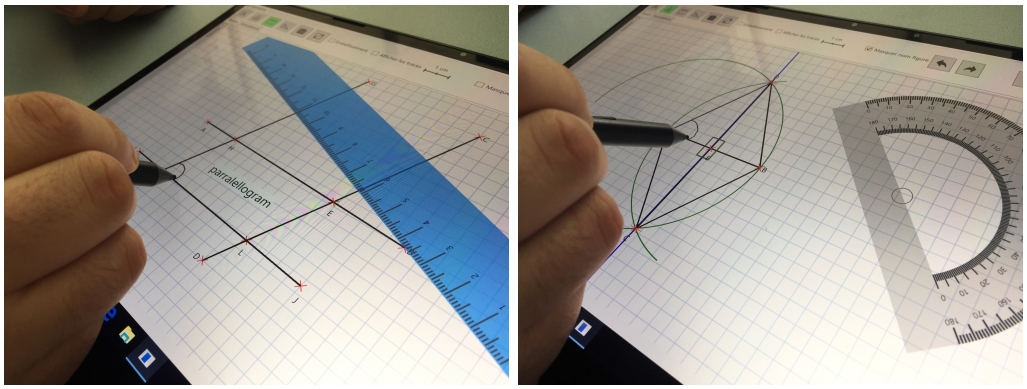


Figure 3: Pen-based and Gesture-based software for geometry learning.

We explore this new research area in collaboration with researchers in psychology of the LP3C/LOUSTIC Laboratory of Rennes. To support this multi-disciplinary challenge, we have developed the new innovative four year national project "ACTIF" (BPI – e-FRAN) with the support of the Brittany Region. The partners for this axe are LP3C and LOUSTIC laboratories, Learn&Go company, educational experts (ESPE) and Brittany region.

2.2.4 Interactive learning of document structure without ground-truth

Interactive Rule Inference We work on the interactive learning of document structure, in the context of a thesis that has just ended. This work enables to combine

[MA09] S. Macé, E. Anquetil, "Eager interpretation of on-line hand-drawn structured documents: The DALI methodology", *Pattern Recognition 42*, 12, 2009, p. 3202 – 3214, New Frontiers in Handwriting Recognition.

statistical methods with syntactical approaches (grammars). Indeed, statistical methods are not able to convey two-dimension hierarchical structures that are common in document analysis. On the opposite, rule-based syntactical methods often require a fastidious manual step for the specification of the various organisations of the document physical layouts. The objective is to model the logical structure with rules and to learn the physical structure. This learning is based on databases of documents with ground-truth that are really costly to label. The current and future work aims at learning physical properties without ground-truth. The scientific context is to lean on large amount of documents and on generic document system analysis. We want to show that some knowledge can stand out from the repetition of physical structures, thanks to non-supervised learning methods. The challenge is to define strategies to make this learning possible thanks to an interaction with the user, which brings a semantic knowledge to the physical detected elements.

Combination Deep Learning / Syntactical Analysis In collaboration with Richard Zanibbi from the Rochester Institute of Technology (RIT), Rochester, New York, USA, we will continue to work on interactive learning by combining deep learning technics, syntactical analysis and user interaction to introduce learning of segmentation. Deep learning methods like convolutional neural networks or recurrent neural networks have shown very interesting results in recognition by being able to make a common segmentation and recognition, with a good introduction of local context. But they are limited to a local context, which is interesting for the recognition of letters and words in a handwritten text line, but is not enough for a modeling and an understanding of a complex structure like the one we can find in a complex structured document. We propose to study the strong combination of deep learning and syntactical methods to build a document structure recognition system able to deal with segmentation problems by learning them. The syntactical part models the structure and brings complex context to the deep learning recognition. The objective is to introduce in the architecture of the neural networks the large contextual information and to make the neural networks able to give not only a recognition but also information of localization of the recognized element. Indeed this localization information is important for the syntactical part to continue and explore different solutions in the global recognition of the document. To train the neural networks, we will have to focus also on a semi–automatic generation of datasets and ground truth, made by the grammatical description of the document, in combination with unsupervised clustering and a user interaction to generate ground truth with a minimum of manual work.

Spread Applications These combinations could open large perspectives by simplifying the grammatical description as much as possible by learning the document structure, including regions of interest (segmentation), region types (classification) and their relationships (parsing/structure). Many applications could be studied on domains where it is important to combine deep learning and strong a priori knowledge. We will also make this combination able to deal with born digital documents (pdf, XML...) to address the huge quantities of documents, which need a real understanding for information extraction.

2.2.5 Document collection analysis for historical big-data

Strategies for Sequential Collection The DMOS-PI method proposes a framework for the analysis of collections of documents. It enables to share information from the collection between the pages, thanks to an iterative mechanism of analysis. This mechanism also makes it possible to integrate an asynchronous interaction between automatic analysis and human operators. We propose to work on modeling strategies of analysis for the analysis of collections of documents. The strategies could sequence the various iterative treatments of documents pages, the global treatments and the interactions. The interest is to exploit as much knowledge as possible on the collection in order to make the extraction of information in each analyzed pages more reliable, and to make the understanding between the various data at the collection level easier. In this context, the ANR HBDEX project has been selected. It is led by the PSE "Paris School of Economics" ("Ecole d'économie de Paris"), with the LITIS lab in Rouen and the CAMS-EHESS. This project focuses on the extraction of historical big-data for digital humanities, applied to financial data. The objective is to analyse masses of tabular data: daily listing on the Stock Exchange from the 19th and 20th centuries. The analysis will be based on the redundancy between the successive days of listing and the consistency between the global sequences of data. This modeling will enable a fast adaptation to other kinds of historical tabular data that only exist on a paper form (economic, demographic, meteorological), but that is necessary to constitute historical big-data databases. This opens a large possibility of applications on documents found in all statistical institutes.

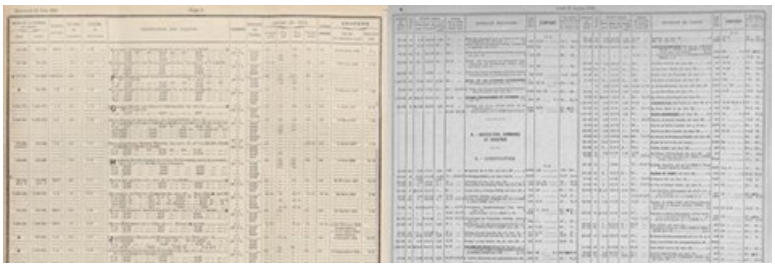


Figure 4: Examples of daily listing on the Paris Stock Exchange

Adaptive Document Layout Analysis We propose to integrate the interactive document structure learning without ground truth and the collection modeling to generate an adaptive document layout analysis system where a user, with few interactions, could make the recognition system learn new layouts to adapt itself and improve the global recognition quality. We will build this adaptive document layout system on the European project EURHISFIRM (InfraDev). EURHISFIRM designs a world-class research infrastructure (RI) to connect, collect, collate, align, and share detailed, reliable, and standardized long-term financial, governance, and geographical data on European companies. This project is leaded by the PSE "Paris School of Economics" ("Ecole d'économie de Paris"), with seven partners working on quantitative economics and finance, economic and social history, and the LITIS Lab in Rouen working with us on document images analysis. We will work on a system to extract high-quality data from historical serial printed sources, to address three issues: (i) lowering the costs of data

extraction from the same source; (ii) lowering the cost of adaptation of the system from one source to the other; (iii) developing effective data validation process. Interactions between the system and experts on the sources lay at the heart of the conception.

2.3 Application domains

[to be removed] Describe in short subsections or as an itemized list the application domains the team is adressing and/or could be adressing. This section does not have to be long.

3 Scientific achievements

3.1 Mathematical expression recognition and analysis

Participants: Arnaud Lods, Éric Anquetil, Sébastien Macé (Learn&Go company).

This part summarises the work done for the end of the third thesis' year extension granted due to delay caused by sanitary restriction put in place in all primary schools during the year 2020. This thesis made by A. Lods began on January 2018 on the subject of On-line analysis of handwritten arithmetic operation on digital tablet: Design of an innovative educational solution to improve learning arithmetic calculations in elementary school. This thesis is carried out in collaboration with the LabCom "Script&Labs" (see section 5.2.4) and is financed as a CIFRE by the ANR n°ANR-16-LVC2-0008-01.

The objective of the thesis is to produce a pen-based tablet application to provide children in elementary school with an application to practice solving arithmetic operation (addition, subtraction, multiplication). Given a mathematical problem proposed by a teacher, the child is expected to solve this problem the same way he would using a pen on a paper. The system then analyzes the student handwritten input and analyse it knowing the expected solution. If errors are detected because of miscalculation, algorithm misunderstanding, omission (if he forgot to carry over in addition or subtraction) or misplacement, the system guides the student to fix his mistakes.

From our previous work we produce from a set of handwritten strokes a set of hypotheses graphs representing the operation. An hypothesis selection was proposed to find a quick best correspondence between the hypotheses graphs and the expected answer graph generated from the instruction. This year contribution has been focused on the hypothesis generation. To avoid creating a large number of unlikely hypothesis, on which we have to complete at least a partial matching, we propose in [7] a new iterative process to speed up the matching and hypothesis generation process. Using the process previously presented, we generate a first 'best' hypothesis given only the segmentation scores. This hypothesis is segmented in a graph of lines for a quick first-match with the expected answer. If any line from the graph is perfectly matched, then the related symbols are considered correct in both graphs. Then we backtrack to the segmentation hypothesis, and for each non-corresponding strokes, we generate new segmentation hypotheses with an increased segmentation threshold. We use the line matching to select and keep the best segmentation hypothesis which will be used for the complete matching. This operation can be repeated until a maximum segmentation

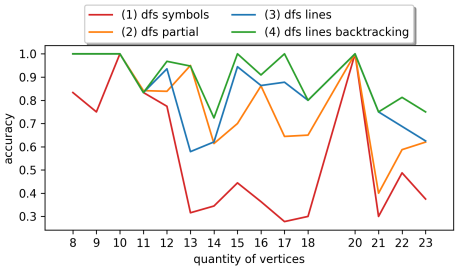


Figure 5: Accuracy for each graph size. The accuracy is computed on the resulting analysis: if dissimilarities found were expected, the analysis is correct

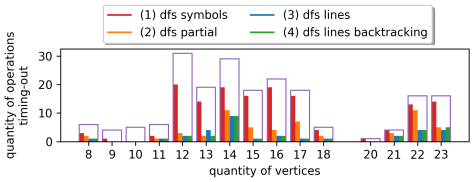


Figure 6: Quantity of operations reaching time-out for each graph size confronted. The purple indicator represents the quantity of operations in the dataset for each graph size.

threshold is reached or when a corresponding match was found for each stroke. If at least one vertex isn't matched after this process, we compute a complete matching on the remaining vertices. Thanks to the backtracking steps, we avoid generating a large number of hypotheses but we are still able to correct early segmentation mistakes.

Figure 5 presents the accuracy score obtained for different sizes of graphs for our previous methods confronted to the new backtracking process. Figure 6 presents the quantity of operations reaching the time-out before the end of the computation for each method. A time-out of 5 seconds is considered. We can see that thanks to the backtracking, we can complete the computation on a larger number of operations, resulting in better accuracy scores as we are able to complete the matching on the correct hypothesis instead of selecting and keeping an imperfect matching.

3.2 An anticipation strategy to manage non stationary data stream: Application to the on-the-fly gesture recognition

Participants: Clement Leroy, Éric Anquetil, Nathalie Girard.

Keywords: Online learning, Non stationary datastream, Evolving fuzzy system.

The following results come from the third PhD years of Clement Leroy entitled: "Incremental learning in non stationary stream: Application to the on-the-fly gesture recognition".

Working with non stationary datastreams require for the classification system to evolve its model over time. In particular, different type of concept drift can occurred in the stream (brutal drift, gradual drift ...), and the model have to apply appropriate adaptation regarding the type of occurring drift. It exists different family of adaptation with features as blind adaptation, adaptation based detection, structural adaptation, parameters adaptation, local adaptation, global adaptation. Each of them tackle different type of concept drift (see Figure 7). However each of these adaptation methods competes with each others making the classification system not optimal and even more unstable when they are combined. On the other hand, the anticipation concept defines in the first two years, and integrated in a new architecture of evolving fuzzy system

namely ParaFIS, have allowed to explore different ways to adapt a same model by making different hypothesis on the evolution of the stream [6]. In this ways, only one particular adaptation method is used to learn one hypothetical system based on one hypothesis done over the occurrence of concept drifts.

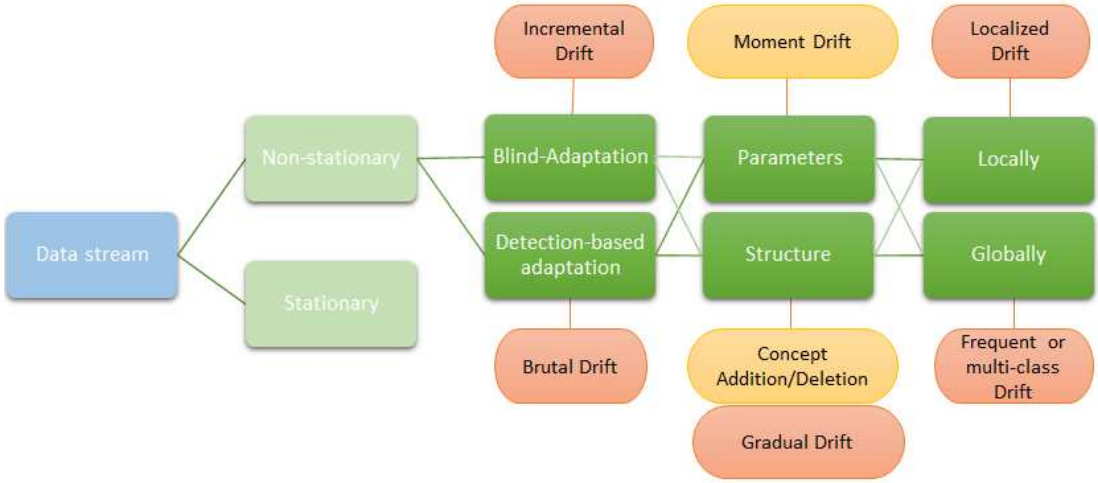


Figure 7: Synthesis of the different adaptation families, and the drift that tackles

During this last PhD year, different strategies to learn ParaFIS were explored. Each strategy makes a different hypothesis on the type of concept drift that can occurred in the stream. Among others, there are ParaFIS without forgetting, ParaFIS with continuous forgetting, ParaFIS with forgetting based on detection, ParaFIS with addition of new rules, ParaFIS with a rule freeze module ... The benefits of each method were evaluated on artificial data sest where the stream is characterised and the drift occurrence is known. This work is in its early stages. In end, the goal is to propose a complete system (see Figure 8) able to adapt all these different types of drift by using all the different strategies in an anticipation module and selecting each time the most appropriate to the current context, and on the basis of a detection measure.

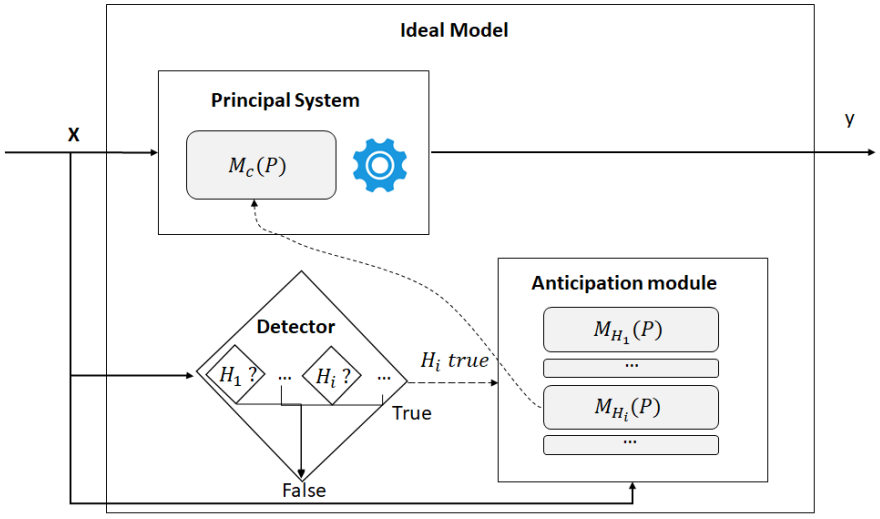


Figure 8: The ideal complete system based on three agents. A principal system, learned incrementally, that classify data, the anticipation module that learned in parallel different version of the principal system using different strategies based on different hypothesis done over the concept drift occurrence, a detector that can test each hypothesis H_i done. Once a hypothesis is verified, the model M_{H_i} in the anticipation module replace the principal system.

3.3 Early recognition of 2D gestures

Participants: William Mocaër, Eric Anquetil, Richard Kulpa.

Keywords: Online Handwritten Gesture Recognition, Convolutional Neural Network.

Funded by the IntuiDoc team and by the EUR Digisport, this thesis has the objective to combine the knowledge of two teams, IntuiDoc and MimeTIC, to solve a common challenge: early gesture recognition.

It can be 2D gestures; symbols drawn on tablets, or 3D gestures: gestures done by an human (hands or full body). The final goal of early recognition is to be able to elaborate very reactive applications. First we focused on early recognition of 2D gestures.

Inspired by 3D gesture recognition approach, we built a new CNN architecture called OLT-C3D (for Online Long-Term Convolutional 3D) based on a spatio-temporal CNN. The originality of the network is it's ability to take a skeleton representation along the time, and apply spatio-temporal convolutions which does not consider the future. Also, thanks to dilated convolutions, the network is able to see far in the past and we can avoid the use of recurrent cells, the network is shown in figure 9.

The representation of the 2D gesture given in input of the network is a sequence of images, which fixed-length increments between each frame.

To make the network take early decisions, we used a confidence mechanism based

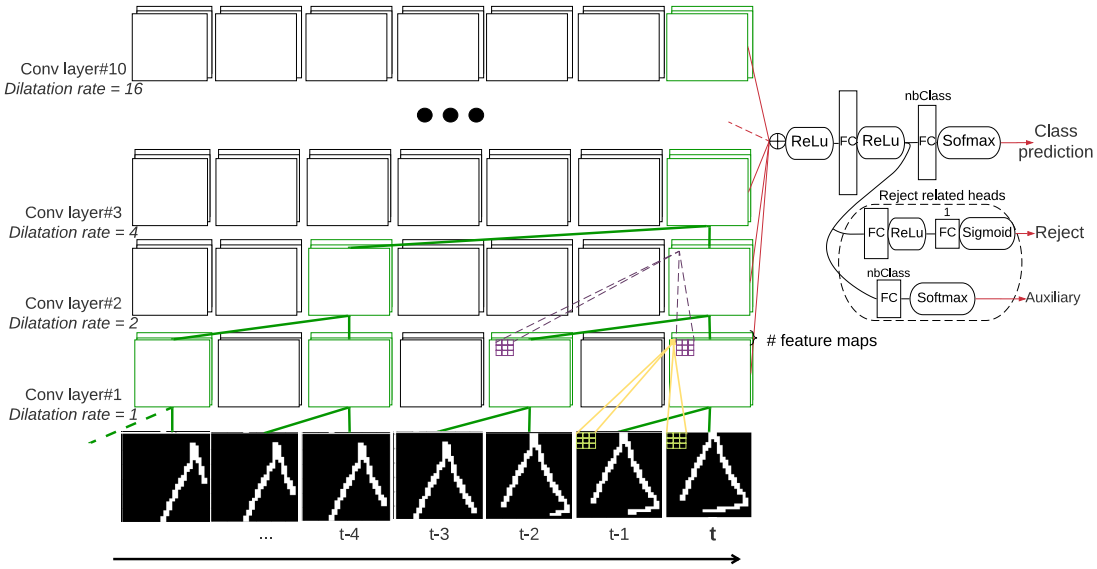


Figure 9: Architecture of the network OLT-C3D designed to handle long-term information, in an online context, using only convolutions.

on SelectiveNet. At each frame we have a confidence score and if it is over a threshold then the prediction is accepted.

This works well for trimmed 2D gesture recognition, and this lead to a publication [8] at ICDAR2021 (oral presentation in September in Lausanne).

3.4 Online Children Handwriting Recognition

Participants: Simon Corbillé, Éric Anquetil, Élisa Fromont (LACODAM Team).

Keywords: Online handwriting recognition, Deep learning, Digital learning, Children handwriting.

This section presents works done in the second PhD year of S. Corbillé which began on 1st October 2019 and entitled: Hybridization of AI approach "Transparent" and "Deep Learning" for the automated analysis of graphic productions of students in the context of education.

The goal of this thesis is to used transparent methods and deep learning approach to do **online children handwriting recognition**. Transparent methods are based on expert knowledge, and the results are understandable by a human while deep learning methods obtain very good result but can have a black box effect so the results can be difficult to understand. The idea is to use them jointly for benefit the advantage of each other.

During the first year, the work was to use online signal with an image approach to do **character recognition**. Our contribution is to add dynamic information contain in online signal which characterise handwriting such as direction and orientation. Theses results have been published in the ICFHR conference.

In the second year, the work focus on **recognition and segmentation of handwriting sentence**. Because of complexity add of sentence level, the previous approach cannot be used. Two architectures were compared for this task: Convolutionnal Recurrent Neural Network (CRNN) and Encoder Decoder model with attention. A Connectionnist Temporal Classifier is used to do the transcription. This type of model achieves good result for the recognition part. A character segmentation can be done with computing receptive fields use for the character prediction (for the CRNN) and the attention maps position (for the encoder decoder). The problem is the quality of segmentation is not good enough to do a feedback on character with good confidence. A post process using a decomposition of handwriting in primitives (chunk of character) can be added to enhance the result but it is not enough again to use it in real application with a good confidence score. Fig. 10 illustrate segmentation examples. A parallel work is to use both deep learning model and a existing system develop in IntuiDoc team which use expert knowledge to do the recognition and segmentation. The idea behind this hybridization is to use the strength of each system. Indeed, the deep learning model achieve better recognition performance and the existing model achieve better segmentation result.

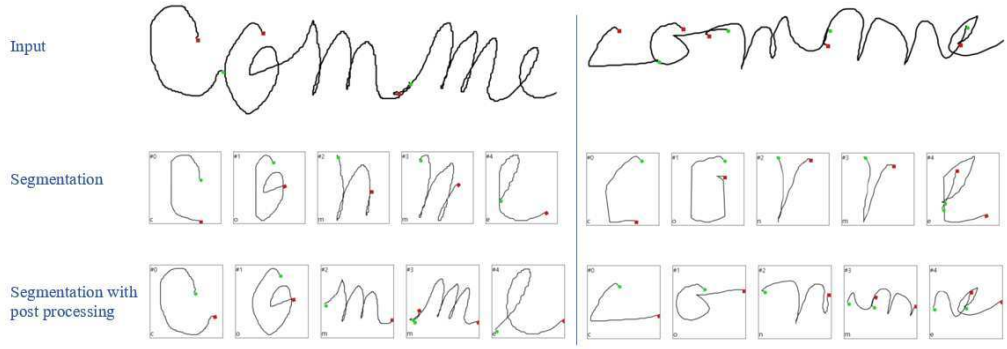


Figure 10: Example of word segmentation

A third part under development is to use a Convolutional Prototype Network and CTC regularization technique to enhance the segmentation result of Deep learning model. The idea is to use a model more precise and confident for character prediction.

3.5 P2IA Cycle 2: Analysis of French spelling in handwriting

Participants: Éric Anquetil, Nathalie Girard, Simon Corbillé, Omar Krichen, Pauline Nerdeux.

Keywords: Online handwriting analysis, spelling analysis.

This section presents research done in the context of the Innovation - Artificial Intelligence Project (P2IA) of the french Ministry of Education (see section 5.2.3) to provide teachers with a teaching assistant based on artificial intelligence for helping to learn French (writing/ spelling). In this project, the objective is to develop an artificial

Figure 11: Feedback examples for dictated words *"mes"*, *"alors"* and *"bien"*

intelligence engine able to analyse the handwritten words of pupils in CP, CE1 and CE2 in a context of dictation exercise, and to provide relevant feedback on spelling errors. A first part of the work have in adapting the word analysis chain, designed for a copying context in the "Kaligo" application, to the dictation context which brings new difficulties. The first difficulty is that the handwriting can be more degraded than in a copying context, since the pupil does not see the word to reproduce. The second one consists in the wide variety of spelling errors a pupil can make in this dictation context. We have designed two strategies to deal with this analysis task:

- A double input strategy asking the child to type the handwritten word;
- A single input, phonetic hypotheses generation strategy.

3.5.1 Recognition with prior knowledge of the handwritten word: Double input strategy

In this first and straightforward strategy, the child types on a keyboard what he or she has written with the stylus. This allows the system to perform better segmentation and identification of handwritten letters, to identify spelling mistakes, and to generate relevant feedback (for example, letters or accent insertion in red and substitution in orange, *c.f.* figure 11). This strategy has led to the delivery of a first prototype to the company Learn&Go in 2020 which was been experimented in several schools in France. An optimized version, with better segmentation and spelling mistakes recognition performances, has been developed and delivered in 2021.

This method is appreciated by some teachers who see a pedagogical interest in training children to type with a keyboard. However, this additional and necessary step to analyse the written word can be seen as a major drawback by other teachers and that is why research was carried out this year to be able to analyse the spelling of a dictated word without this prior knowledge.

3.5.2 Recognition without prior knowledge of the handwritten word

The second strategy aims to be free from the user defined groundtruth and to predict it given only the instruction and the handwritten strokes. It is based on the integration of a phonetic hypotheses generation module to the analysis workflow. This module is based on the **Phonetisaurus engine**, a stochastic Grapheme to Phoneme (G2P) WFST (Weighted Finite State Transducer). For a given instruction, this engine generate a set of phonetically similar pseudo-words. Since the ground truth is unknown, the new analysis chain is guided by all the generated phonetic hypotheses from the instruction

(dictated word). Table 1 presents the performance of each analysis mode/strategy on the test set in terms of correct segmentation and recognition. The fact that the ground truth is already available in the analysis with double input mode allows this strategy to have the best analysis rate (80.6 %). The error rate of 19.4% demonstrates the complexity of the task in hand. The phonetic analysis strategy achieves a lower recognition rate (69.4 %), however by far better than the existent copying analysis strategy (with only the instruction as prior knowledge). This work has been the object of a publication at GREC (ICDAR Workshop on Graphics Recognition) [4].

3.6 LabCom Script&Labs: Fine Handwriting Quality Analysis: going further with handwriting

Participants: Éric Anquetil, Nathalie Girard, Omar Krichen, Pauline Nerdeux.

Keywords: Online handwriting analysis, character analysis, word analysis.

The work done in 2021 in the scope of the P2IA project has had a positive impact on the analysis performance of the word analysis engines developed in the previous years in the scope of the Labcom Script&Labs.

As the LabCom ends at the end of the year, a delivery of the source code related to the handwriting analysis engines developed in the scope of this partnership has been done to the company Learn&Go in July 2021.

3.7 Kaligo Luxembourg

Participants: Eric Anquetil, Nathalie Girard, Omar Krichen, Pauline Nerdeux.

Keywords: Online handwriting analysis, character analysis, German handwriting.

After having developed the handwriting analysis engines for the French and English languages for the company Learn&Go within the scope of the Labcom Script&Labs (see section 5.2.4), we are now developing the analysis engines for numbers, upper and lower case letters for the German language in order to allow the company Learn&Go to release their Kaligo application in Luxembourg.

In order to manage this new language, work has been done to allow the analysis of new letter forms such as the Eszett letter (ß) or lower and upper case umlaut characters (ä, ö, ü, Ä, Ö, Ü).

Table 1: Analysis performance of each strategy

Approach	Correctly analysed	Analysis rate
Analysis guided by the instruction	633	58,72%
Double input analysis (childtyping)	869	80.6%
Phonetic analysis	748	69.4%

A first delivery of the lowercase letter analysis engine and of the digits analysis engine has been done on December 20th, 2021 and the delivery of the uppercase letter analysis engine is planned at the beggining of the year 2022.

3.8 e-Fran national project: Real-time interpretation of geometric shapes for digital learning

Participants: Omar Krichen, Eric Anquetil, Nathalie Girard.

Keywords: Online recognition of Hand-drawn sketches, Intelligent Tutoring Systems, planning.

This work is the result of the PhD of Omar Krichen in the context of the « e-Fran » national project called ACTIF (see section 5.2.5) and deals with the design of pen-based intelligent tutoring system for geometry learning in middle school, called IntuiGeo. This tutor simulates the traditional pen and paper approach, such as the pupil can draw freely with the stylus while manipulating, with his fingers, virtual tools, such as a ruler or a compass. The objectives of this work is to ensure the pupil's learning transferability between the digital support and the traditional one, and to provide adapted assistance to improve the pupils performance. This tutoring system is composed of two main engines: a recognition engine which interprets on the fly the user's strokes, and a supervision engine which is responsible of the tutoring aspect of the system. This supervision engine is composed of a domain module, encompassing the declarative knowledge, a learner module, which interprets the pupil's actions, and an expert module, which encompasses the procedural knowledge of the domain. This year has been focused on the development of an author mode, which enables the intuitive creation of new construction geometry exercises. In this mode, the teacher only has to sketch a solution example for the system to generate the problem automatically (constraints-based modelling with a recognition graph). The teacher also has the possibility to propose a partially completed geometric figure, which the pupil has to complete (this is a typical exercise in geometry). Another interesting point is the ability to parameterize the using of virtual tools (compass, ruler...), by forbidding the usage of certain ones for a given exercise. The system verifies the validity of these teacher defined constraints, by trying to synthesize strategies to solve the problem without using the forbidden tools. A publication has been submitted to Educational Technology Research and Development, which relates experiments that have been conducted in classrooms in Brittany, and demonstrate the positive pedagogical impact of the system. The publication is currently under revision.

3.9 Interactive and iterative analysis of price-lists

Participants: Camille Guerry, Simon Bouvier, Iwan Le Floch, Bertrand Coüasnon, Aurélie Lemaitre.

This work is done in the context of the ANR project HBDEX (see part 6.2.5). The aim is to model a strategy that automatically exploits the sequential aspect of a collection to improve document recognition.

The strategy that we propose is composed of different iterations. The aim of each iteration is to recognize and validate one kind of structural element in the documents at different levels of granularity. An iteration consists of:

1. a first individual structural analysis of each document, based on a combination of deep-learning and syntactical approach,

2. a transverse validation phase, using knowledge coming from the context of the collection,

3. a second individual structural analysis. For this step, the recognition system exploits the information validated by step 2 to specify the grammatical description used in step 1.

For the validation phase (step 2.), we first construct a sequence composed of data that comes from the structural analysis phase (step 1). We then apply a validation process to this sequence. We proposed two different validation processes, depending on the knowledge that we have one the collection.

- Rupture detection: we apply a bottom-up rupture detection method used in signal processing [TOV20] to detect changes in our data. We then post process the obtained slots to detect recognition errors and correct them.

- Sequence alignment: we proposed a method inspired by sequence alignment methods used in bioinformatics to identify elements coming from different images.

After this validation phase, we generate questions for the user if there are ambiguities or if the information needed is not available in the original images.

In a first instance, we focus on daily price lists coming from Paris stock exchange markets of the end of the 19th and beginning of the 20th century. This sequential collection is composed of 40 years of documents (more than 40 000 images).

We evaluate our strategy on column localisation and stock identification tasks. Our method significantly improves column localisation (from 7.89% error rate to 0.44% error rate) and stock identification (from 0.914 to 0.988 F-measure). Our strategy mainly reduces the number of user interactions required to obtain satisfactory results (from 4,061 to 309).

We also proposed a processing chain that allows this strategy to be easily parallelized to produce a large amount of data. We are currently producing data for daily price lists coming from Paris stock exchange markets of the end of the 19th and beginning of the 20th century ("La Coulisse"). In the context of the project Eurhisfirm (see 3.10), we have also tested this strategy on other European price list collections. This strategy is designed to be generic and could be applied to other types of collections.

[TOV20] C. Truong, L. Oudre, N. Vayatis, "Selective review of offline change point detection methods.", *in : Signal Processing, 167, 107299*, IEEE, p. 770–778, 2020.

3.10 Analysis and recognition of price lists

Participants: Simon Bouvier, Iwan Le Floch, Jean Camillerapp, Bertrand Coüasnon, Aurélie Lemaitre, Yann Ricquebourg, Nathalie Girard.

This section presents the work developed as part of the EURHISFIRM project (see section 5.1.2). The goal is to be able to extract long term data on European companies, from scanned 20th century price lists. We worked on documents from Paris, Bruxelles and Madrid, which possess different structures.

Last year, we began to make a recursive and generic system able to segment and order all the tables present on a price list using structural clues. This system describes a global meta table structure whatever the number of rows, whatever the number of columns it is made of, and tries to detect recursively in each detected cell, another recursive table structure. This description is done using a bi-dimensional grammatical description of table structures and is able to detect the global meta table structure in any price lists tables. The system can use double or thick vertical or horizontal line borders, and/or understand the recursive organization of the table to detect the global table. In case documents are degraded or if a line segment representing a line border is damaged, the system is also able to correctly detect the global table structure. We can see on Fig. 12 examples of meta tables, on pages from Bruxelles 1931. Here, each table possesses a pair of two numbers :

- The first one represents the type of table, which is determined by the header and columns. It is useful in order to identify consecutive tables.
- The second one represents the reading order inside a same type of table

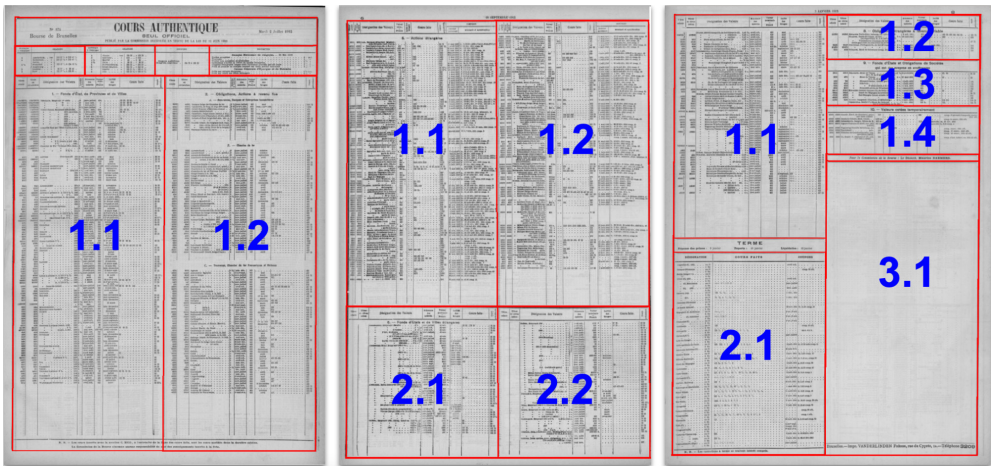


Figure 12: Example of meta tables from Bruxelles 1931 price lists.

Once we were able to segment and order tables, we used in each table the data extraction method developed by Camille Guerry as part of the ANR project HBDEX (see section 5.2.6). This method was made to work on one specific type of price list, so we updated it, to make it generic and easily adaptable to new structures and price lists. Our system always analyses a document the same way, but we specify characteristics

that slightly modifies the process when needed. That way, each type of document is characterized by a list of attributes and not only by their origin. These attributes could be used again in future price lists documents, to produce in an easy way an adapted version combining in a different way the different characteristics. Here, we can see an example of results obtained with our system on figure 13.

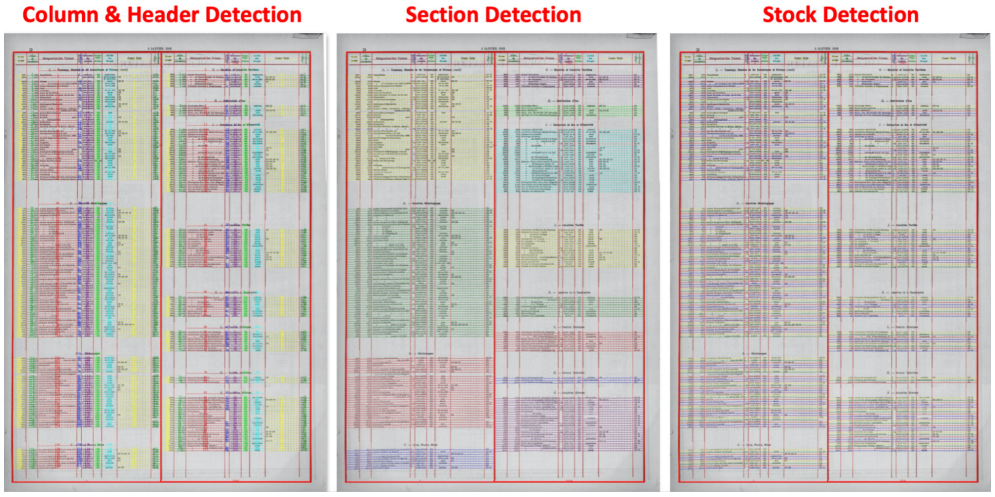


Figure 13: Example of detections on Bruxelles 1931 price lists.

3.11 Transformer-Based Neural Network for Handwritten Text Recognition

Participants: Killian Barrère, Bertrand Coüasnon, Aurélie Lemaitre, Yann Soullard.

Keywords: Transformer, Handwritten text recognition, Neural networks.

Here, we summarize the works achieved by Killian Barrère during the first year of his thesis entitled "Deep neural networks and attention mechanisms for handwritten text recognition". The thesis aims to study recent innovations in neural networks and their application to the field of handwritten text recognition.

First works were dedicated to experiments related to encoder-decoder (or sequence-to-sequence) models for handwritten text recognition. Then, with the increasing interest for transformer models in various fields related to neural networks, we performed intensive experiments related to a transformer-based architecture for handwritten text recognition.

Training a transformer-based architecture for handwritten text recognition represents a challenging task due to the relatively limited number of annotated data. Therefore, we proposed a light transformer-based architecture. The architecture remains low in the number of parameters and might perform well even with few annotated data. We used an encoder based on both convolutional layers and self multi-head attention, and a decoder composed of self multi-head attention and mutual (or encoder-decoder) multi-head attention. To train efficiently the proposed architecture, we use a hybrid

loss combining a connectionist temporal classification loss with a cross entropy loss. The encoder is trained with the first loss function, while the decoder is trained with the second.

We tested our architecture on modern English handwritten text and obtained results close to the state-of-the-art. The architecture we proposed has less than 10M parameters and is much smaller compared to other transformer-based architecture, which might use up to 100M parameters. With the proposed architecture, we are able to obtain results at the level of other transformer-based architectures, without additional data. Such additional data are crucial for some architectures.

We presented the results during Doctoral Consortium of the 2021 International Conference on Document Analysis and Recognition (ICDAR) which resulted in the best poster award and a publication [3].

3.12 Analysis and recognition of historical parish registers

Participants: Solène Tarride, Bertrand Coüasnon, Aurélie Lemaitre, Jean Camillerapp, Ivan Leplumey.

Keywords: Handwritten text recognition, named entity recognition, neural networks, historical documents.

This work is carried out in the context of a CIFRE PhD (n°2018/0896) in collaboration with Doptim (see section 5.3.2). The aim of the project is to develop strategies for automatic recognition of historical handwritten documents [2]. More specifically, we focus on French demographics documents, in which baptisms, marriages and burials were recorded.

We have explored sequence-to-sequence neural networks with attention mechanisms for handwritten text recognition (HTR) and Names Entity Recognition (NER). Our aim is to assign to each recognized word a semantic category, such as *name*, *family name*, *place*, *date*, or *occupation*. We have explored two strategies for this task, as illustrated in figure 14. *Sequential strategies* tackle handwriting recognition before named entity recognition. *Joint strategies* tackle handwriting and named entity recognition at the same time. In this scenario, the network predicts characters as well as semantic tags to localize relevant words.

We have compared these two training strategies using the same seq2seq architecture on the Esposalles public database. Our results highlight the interest of joint approaches strategies for information extraction, as text recognition seems to benefit from semantic knowledge. Moreover, sequential strategies tend to propagate errors: inaccuracies introduced during the transcription phase can lead to misclassification of named entities. Finally, we have also explored multi-source and multi-task training strategies for joint HTR and NER using sequence-to-sequence networks. Results indicate that attention-based models benefit from tackling multiple tasks at the same time. Our study have been submitted in the 2022 International Workshop on Document Analysis Systems (DAS).

(a) Sequential strategy

(b) Joint strategy

Figure 14: Two strategies for HTR and NER. Semantic categories: occupation of the wife's father (blue), location of the wife's father (yellow), name of the wife's mother (pink).

3.13 Generating realistic synthetic historical documents

Participants: Killian Barrère, Solène Tarride, Bertrand Coüasnon, Aurélie Lemaitre, Yann Soullard.

Ongoing works are dedicated to the production of realistic synthetic data of handwritten text. Synthetic data might help existing neural networks to perform better. This is especially valuable for handwritten text recognition where annotated data are expensive. It will be even more meaningful as we also plan to apply transformer-based architecture to historical documents, with even less annotated examples.

We have designed a pipeline for synthetic document generation. The three steps of this pipeline are illustrated in figure 15. First, the text appearing in the document image must be generated. It can either be extracted from Wikipedia (for modern text of any language), historical books (for specific language and time period), or generated from template-based models (for semi-structured documents). Then, the image is generated using handwritten fonts that are selected at random for each document. Finally, image transformations, such as elastic transforms, baseline curving and ink fading, are applied to make the document more realistic. The pipeline is integrated in a Pytorch dataloader, and can be easily used for on-the-fly data generation during training of any neural network.

The work on synthetic data is a collaboration between Killian Barrère and Solène Tarride and will be carried on during the following months.

3.14 Unsupervised learning of face image representations

Participants: Martin Dornier, Bertrand Coüasnon, Christian Raymond, Yann Ricquebourg.

Keywords: Neural networks, Unsupervised learning, facial landmark detection.

This CIFRE PhD (n°2020/0133) in partnership this InterDigital R&D France focuses on training neural networks with minimal annotated data in several InterDigital's fields of area.

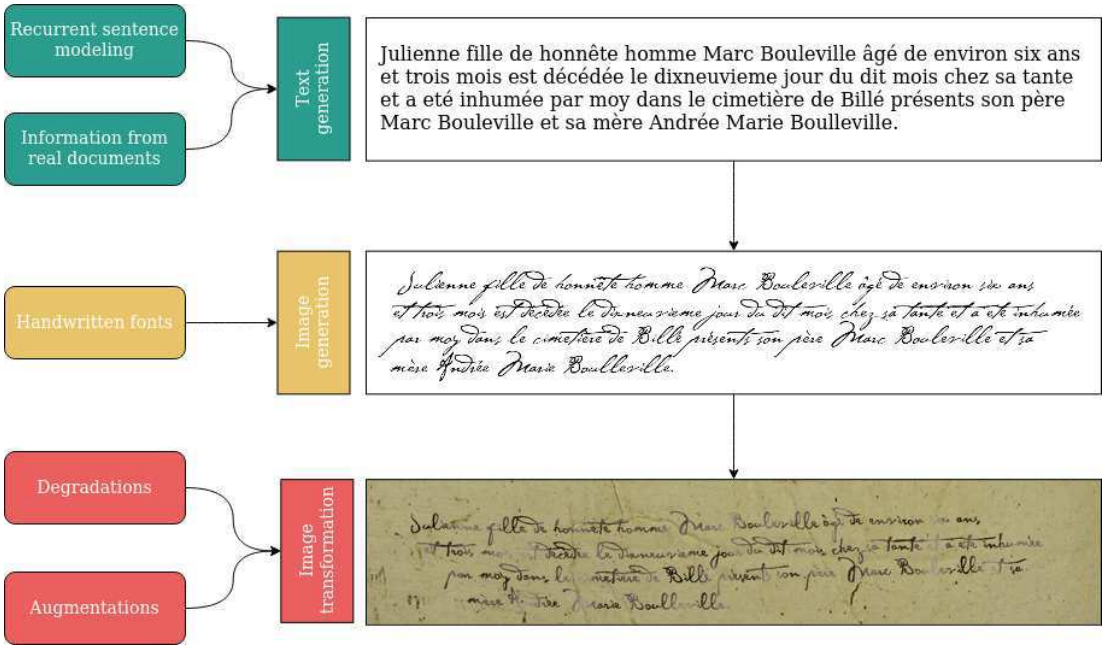


Figure 15: Illustration of the pipeline designed for synthetic document generation. In this example, synthetic french demographic records are generated. The text is generated using a template-based model of French death record. Real names, surnames and locations are used to fill the template.

During the first year of this PhD we focused on facial landmark detection with few annotated data. A first approach based on a bi-directional GAN has been tested but it did not lead to successful results.

A second approach was based on a published paper has been developed. The article architecture based on an auto-encoder has been enhanced with the addition of "skip-connections". An active learning scheme has also been implemented. Thanks to these contributions, the number of annotated training samples needed for the facial landmark detection training has been greatly reduced.

A article based on this work has been submitted and accepted to the International Conference on Image Analysis and Processing (ICIAP) 2021. Due to the Covid situation, this conference has been postponed to May 2022.

3.15 Historical Map Segmentation

Participants: Aurélie Lemaitre, Jean Camillerapp.

We took part on the competition of Historical Map Segmentation, in ICDAR'21. The goal is to segment document images of Paris maps from the beginning of the 20th century: delineate the content of the map and locate the graticule line intersections, which give indications on the geographic coordinates of the map.

Our contribution is entirely led by a rule-based method. Thus, it does not require a training phase, nor annotated data. We extract the line segments, in the image,

at various resolutions. Then, a grammatical description combines those elements in a logical way to filter the relevant map contents. This is an original approach, at the era of deep learning techniques, but that is very convenient for old non-annotated documents.

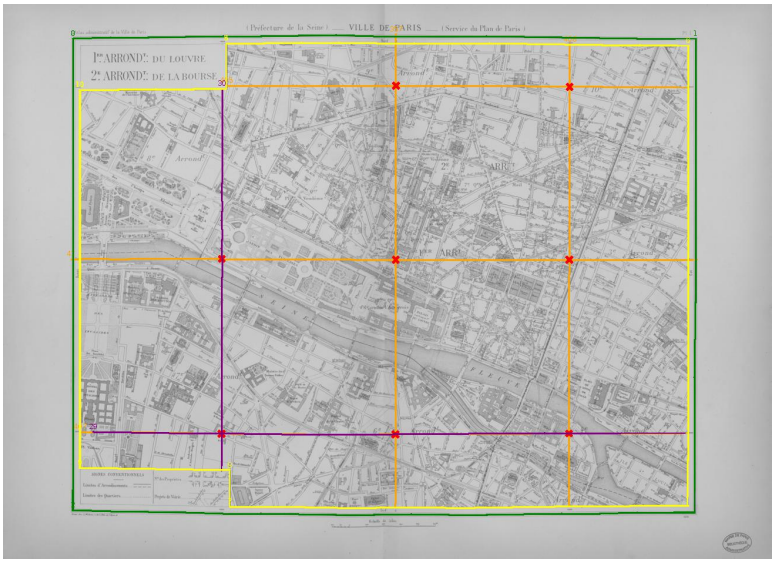


Figure 16: Result of map segmentation: in green and yellow the border of the map; In purple the reference cross for graticule lines, used as a base to detect the rest of the grid in orange. In red: intersection of graticule lines.

We validated this work during MapSeg competition (figure 16). It obtained the third position for map segmentation and the second position for graduate line detection, with a localization score of 89.2%. The results are published in [5].

4 Software development

4.1 Software Deposit

All the presented softwares have been deposited in APP. More details on those softwares can be found on Intuidoc web site (http://www.irisa.fr/intuidoc).

4.2 IA for Digital learning: Handwriting analysis software

Contact: Eric Anquetil

Keywords: Handwriting Analysis, digital learning, fuzzy logic.

In the area of digital learning associated to the IntuiScript project and the ANR LabCom ScriptAndLabs (see section 5.2.4 and 3.6), we have developed four software:

- The ISF (Isolated Symbol Feature) software library allows the characterization of the meaning and direction of on-line manuscript tracing.

- The ISA (Isolated Symbol Analysis) software library allows the analysis of isolated symbols: capital letters, cursive letters, numbers
- The IWA (Isolated Word Analysis) software library enables the analysis and segmentation of handwritten words.
- The DAP (Drawing Precision Analysis) software library allows the analysis of the accuracy of a graphical plot against a guidance.

Through industrial collaboration with Learn&Go company, these software have been successfully integrated in the pen-based tablet solution: Kaligo. This solution is distributed by Learn&Go company. It is focused on learning writing at school from children aged 3 to 7. In 2021, at the LabCom Script&Labs end, a delivery of the source code related to these libraries has been done to the company Learn&Go.

4.3 DALI: a framework for the design of pen-based document sketching systems

Contact: Eric Anquetil

Keywords: Sketch recognition, pen-based interaction, visual language theory, industrial transfer.

DALI is a framework for the interpretation of hand-drawn sketches drawn on tablet PCs. The first property of the *DALI* method is its genericity, which means that it can be used to design pen-based software to sketch various natures of documents. It is based on the visual language and grammar theory that makes it possible to model bidimensional symbols and documents *DALI* interprets the user strokes *on-the-fly*, directly during the design of the document; it means that each time the user draws a stroke, the system analyses it and produces a visual feedback, showing how it is interpreted.

This way, the user is an actor of the interpretation process, because he can progressively correct the errors of the system. Thus, the interpretation process can rely on the information given by the user to better interpret the following strokes. The coupling of these two properties increases significantly the efficiency and the robustness of the sketch interpretation process.

The *DALI* method has been used to design several pen-based prototypes, for instance for the sketching of musical scores, electrical sketches, UML class diagrams, architectural floor plans, etc.

It has been transferred to the Script&Go society, which led to the design of *Script&Go Electrical Sketches*. These softwares are today commercialized and used daily by hundreds of technicians in France. *Script&Go Electrical Sketches* has been rewarded with the "Trophées de l'innovation" 2008 for uses, applications and communicating solutions for enterprises", in the category named "Solutions Métiers".

In 2018, DALI framework has been extended to design Dplan (Dali plan) library. The DPlan library allows the analysis and interpretation in real time of pen-based plan sketching on numeric tablet (walls, rooms, doors, windows...). Dplan library has been

integrated in 2018 in the "IntuiDiag" software transferred to the Innax company. This transfer has been supported by a development fund from *SATT Ouest Valorisation*.

In 2021, DAlI framework has been extended to design the DGeo (Dali Geometry) library. The DGeo library allows on the fly analysis of geometric sketches on pen-based tablets. This library has been integrated into the "IntuiGeo" software.

4.4 Intelligent tutoring system for geometry learning

Contact: Eric Anquetil

Keywords: Structured sketches interpretation, intelligent tutoring, bi-dimensional grammars, pen-based interaction.

IntuiGeo is a digital learning software for geometry and has been developed in the context of the efran ACTIF project (see section 3.8). In this context, we have developed these libraries:

- The DGEO (DALI for Geometry) software library allows the analysis and interpretation in real-time of hand-drawn geometric sketches.
- The ITGEO (Intelligent Tutoring for geometry) software library allows the creation of new geometry problems by the teacher, and the real-time analysis and guidance of the pupil's actions in the context of geometry problem resolution.
- The VTGEO (Virtual Tools for geometry) software library enables the manipulation of realistic virtual tools (protractor, compass, ruler, try-square).

These libraries are integrated into the pen-based solution IntuiGeo,which has been tested in middle-schools through the work done in 3.8.

4.5 DocRead : an automatic generator of recognition systems on structured documents

Contact: Bertrand Coüasnon

Keywords: Recognition, structured document, musical scores, mathematical formulae, table structures, forms, archives.

DocRead is an automatic generator of recognition systems on structured documents. It has been developed thanks to the DMOS-P method [Coü06]. It is made of a compiler of the EPF language (with which it is possible to describe a document), a parser associated to this language, an early vision module (binarization and line segments detection) and a classifier having also a reject option.

This generator allows us a fast adaptation to a new type of document. Indeed, it is only necessary to define a new grammar in EPF, which describes the new type of

[Coü06] B. Coüasnon, "DMOS, a Generic Document Recognition Method: Application to Table Structure Analysis in a General and in a Specific Way", *International Journal on Document Analysis and Recognition, IJDAR 8*, 2-3, June 2006, p. 111–122.

document. Then, if necessary, a new learning of the classifier is done to make it able to recognize new symbols. The new recognition system adapted to a new structured document is produced by compilation.

With this generator, we already have been able to produce recognition systems of structured documents. The recent, years, we mainly focused on the following ones:

- TabRead: a prototype for table structures recognition;
- Mexicanread: a software for field localisation in ancient mexican marriage records [LCCC18].
- MapRead: a software for the segmentation of historical maps [5].

4.6 DocRead development platform

Participants: Jean Camillerapp, Aurélie Lemaitre.

As presented in the previous section, DocRead is a method to develop new recognition systems. We set up a complete development environment for the creation of recognition system. This environment is based on:

- a virtual machine containing all the appropriate libraries,
- a kernel of lambda Prolog, PM-MALI,
- a complete environment with Eclipse, with a dedicated plugin for both lambda Prolog and EPF development.

Until 2020, we were using a virtual machine in a 32-bit environment. In 2021, Jean Camillerapp has ported the PM-MALI kernel to a 64-bit environment. This has required a deep study of the architecture of PM-MALI, with many adaptations.

Thanks to this work on PM-MALI, we updated the DocRead development platform with a new virtual machine in 64-bit environment.

5 Contracts and collaborations

5.1 International Initiatives

5.1.1 French-German bilateral ANR project in artificial intelligence (KIHT - Kaligo-based Intelligent Handwriting Teacher)

Participant: Eric Anquetil, Yann Soulard, Romain Tavenard.

[LCCC18] A. Lemaitre, J. Camillerapp, C. Carton, B. B. Coüasnon, "A combined strategy of analysis for the localization of heterogeneous form fields in ancient pre-printed records", *International Journal on Document Analysis and Recognition 21(4)*, 269-282, July 2018, https://hal.inria.fr/hal-01858192.

- Partners: STABILO International GmbH, Karlsruher Institut für Technologie Institut für Technik der Informations¬verarbeitung, LearnAndGo company, Institut National des Sciences Appliquées de Rennes (IntuiDoc, IRISA Laboratory).
- 36 months (2021-2024)
- Contract: INSA, KIHT

In this project, we will design a new intelligent device to help learning handwriting in classrooms. The originality of the project consists in designing a new handwriting capture device developed by the company STABILO: a digital pen equipped with kinematic sensors that allows writing on any surface (screen and paper).

The Stabilo company, supported by the German laboratory KIT, has the task of designing the hardware of the digital pen as well as embedding the AI algorithms developed. On our side, we are working, through a thesis and a post-doc, on the design of an original and powerful deep neural network architecture to automatically synthesise the online handwriting from the kinematic signals produced by the digital pen sensors.

5.1.2 H2020 InfraDev EURHISFIRM: Historical high-quality company-level data for Europe

Participant: Bertrand Coüasnon, Aurélie Lemaitre, Nathalie Girard, Simon Bouvier, Iwan Le Floch, Yann Ricquebourg.

- Partners: Paris School of Economics (PSE)(Coordinator), Universiteit Antwerpen, Johann Wolfgang Goethe Universitat Frankfurt am Main, Erasmus Universiteit Rotterdam, Uniwersytet Ekonomiczny we Wroclawiu, The Queen's University of Belfast, Koninklijke Nederlandse Akademie van Wetenschappen – Knaw, Universidad Carlos III de Madrid, Université de Rouen Normandie, Institut National des Sciences Appliquées de Rennes, Gesis Leibniz-Institut Fur Sozialwissenschaften
- 36 months (2018-2021)
- Contract: INSA, H2020 InfraDev

EURHISFIRM designs a world-class research infrastructure (RI) to connect, collect, collate, align, and share detailed, reliable, and standardized long-term financial, governance, and geographical data on European companies.

EURHISFIRM enables researchers, policymakers, and other stakeholders to develop and evaluate effective strategies to promote investment, economic growth and job creation. The RI provides the tools for long-term analysis highlighting the dynamics of the past and the way those dynamics structure our present and future.

EURHISFIRM develops innovative models and technologies to spark a "Big data" revolution in historical social sciences and valorize Europe's cultural heritage.

In this project we work on building a system to extract high-quality data from historical serial printed sources, to address three issues: (i) lowering the costs of data extraction from the same source; (ii) lowering the cost of adaptation of the system from one source to the other; (iii) developing effective data validation process. Interactions between the system and experts on the sources lay at the heart of the conception. We first started to work on the structure recognition of companies yearbooks.

Current activities are described in section 3.10.

5.2 National Initiatives

5.2.1 ANR CollabScore: Shared spaces for digital music scores

Participant: Bertrand Coüasnon, Aurélie Lemaitre, Yann Soullard, Kwon-Young Choi.

- Partners: Cnam, INSA, BnF, Antescofo, IReMus, Fondation Royaumont
- 48 months (2020-2024)
- Contract: INSA, ANR

The project is dedicated to the collaborative digitization of music scores currently available only as images in museum collections. It will combine OMR (Optical Music Recognition) and a crowdsourcing correction phase of remaining recognition errors. A reconciliation step of the different versions will be automatised with specifically developed software, to obtain a reference score. This fulcrum notation will then be used in conjunction with several sources, to enhance the user experience. For instance, listening could be assisted by the synchronized scrolling of the score, and augmented with musicological annotations. The project aims at solving some scientific challenge, first by guiding and controlling an OMR process with musical knowledge, then by elaborating an automated crowdsourcing process.

5.2.2 Directory of Musical Writings of the Music Department of the BnF

Participant: Bertrand Coüasnon, Aurélie Lemaitre.

- Partners: Université de La Rochelle, INSA, BnF
- 48 months (2020-2024)
- Contract: INSA

Collaboration project on the constitution of a directory of musical writings of the collections of the Music Department of the BnF. Taking into account both autograph manuscripts and manuscripts by identified or anonymous copyists, the project will study the indexing of scripts on graphic characteristics.

5.2.3 P2IA: Project of Ministry of Education (DNE): Project of Innovation - Intelligence Artificial (P2IA) for e-eduction

Participant: Eric Anquetil, Nathalie Girard, Pauline Nerdeux, Simon Corbillé.

- Partner: *Company Learn&Go*, *LP3C*, *Loustic*, *INSPE*, *IRISA-Expression*, *Academies of Rennes and Caen*
- 18 + 24 months (2019-2021).
- Contract: INSA

This project was launched by the Ministry of Education (DNE) as part of the project of Innovation - Intelligence Artificial (P2IA). It is a project funded on several instalments; the first one (12+6 months) is founded with an amount of 240K€ for the IntuiDoc team. The project is led by the Learn&Go company. The objective is to design a software environment for helping to learn French (writing / spelling) based on artificial intelligence for teachers and pupils in cycle 2. In this project, we are responsible for designing and developing the automated analysis engine of handwriting to identify the spelling/graphemes mistakes of pupils in CP, CE1 and CE2.

5.2.4 Script&Labs/ANR joint Laboratory (LabCom): On-line handwriting and drawing recognition and interpretation for active Learning in e-education

Participant: Eric Anquetil, Nathalie Girard, Damien Simonnet, Pauline Nerdeux, Sébastien Thomas, Morgane Carry.

- Partner: *Société Learn&Go*
- 36+12 months (2017-2021).
- Contract: INSA
- Website: `https://scriptandlabs.irisa.fr/`

With the great success of the IntuiScript Project (BPI/PIA-2) (`https://www-intuidoc.irisa.fr/projet-intuiscript/`), we want to consolidate our implication in e-education to achieve innovative contribution based on pen-based tablet devices. In that way, we build a ANR joint laboratory (LabCom) between IntuiDoc and Learn&Go Company on this promising topic for a total amount of 1 562K€(6.15 ETP) with 300K€of grant.

This new structured partnership would have the goal to design new educational learning experience at school by using tablet and tactile digital devices: handwriting learning, generative drawing/sketching, on-line handwritten arithmetic expression, collaborative interaction using 2D gestures, document annotation...

The ScriptAndLabs Laboratory has been selected (from the 122 national LabCom) for an oral presentation of its activities for the "National LabCom days" in Paris in September 2018.

Following this presentation, we have been invited by the ANR President and CEO to present the ScriptAndLabs LabCom at the administration council of ANR the November 22, 2018.

The ScriptAndLabs Laboratory has presented its first research results by an inauguration on October 18, 2018 at the "Jacobin Convent" Palais des congrès in Rennes, as part of the "Learning Show" event.

The developments resulting from Labcom have enabled the commercialisation of KALIGO in France, England and soon in Luxembourg and Germany with high value-added features in 2021. The effective work of the INTUIDOC team has been fundamental to the technical evolution of the KALIGO product.

5.2.5 Actif: Active Learning and Collaboration with Tablet Computer, Interactions and Feedbacks (e-Fran - National Innovative BPI Project)

Participant: Eric Anquetil, Nathalie Girard, Simon Corbillé, Omar Krichen, Morgane Carry.

- Partners: *LP3C*, *Loustic*, *Société Learn&Go*, *Saooti*, *Région Bretagne*, *Espe*, *espace des sciences*
- 48 months (2017-2021).
- Contract: INSA

The project "ACTIF" is one of the 22 selected national project from the "e-fran / innovative national project " call. This is a 4 years project (2017-2020) for a total amount of 1 143 856€, led by the LP3C of the University of Rennes 2. It carries on: "Active Learning and Collaboration with Tablet Computer, Interactions and Feedbacks".

One of the key topics in this project is to investigate how to encourage new learning strategy based on “active learning” and especially on “generative drawing” using pen-based numerical devices. The concept is to ask students to create drawings while reading text for causing generative processing that leads to better learning outcomes. The goal is to improve the learning of student by considering “learning” as a generative activity especially with collaborative/cooperative interaction. In the scope of digital learning, the potential induces by pen-based devices (tablet, TNI and new interactive screen) is really interesting. The goal is to investigate how we can automatically generate intelligent “corrective” or “predictive” individual and collective feedbacks to users by interpreting their drawing process: summarizing, mapping, sketching...

The IntuiDoc team will ensure all the design related to the analysis of graphic productions with one engineer, one postdoctoral researcher and one PhD student. They will work on the issues of "Generative Drawing" in the field of Geometry at secondary school.

We have presented the first research results of the e-Fran projects on January 30, 2019 in Rennes in the presence of Mrs Le Pellec Muller, Rector of the Brittany academic region.

Current activities are described in section 3.8.

5.2.6 ANR HBDEX : Exploitation of Historical Big Data for Digital Humanities

Participant: Bertrand Coüasnon, Aurélie Lemaitre, Camille Guerry, Iwan Le Floch.

- Partners: Paris School of Economics (PSE), Université de Rouen Normandie (LITIS), CAMS-EHESS
- 48 months (2017-2021)
- Contract: INSA, ANR

This project focuses on massive historical data extraction for digital humanities, applied to financial data. The objective is to analyze masses of printed tabular data: daily stock exchange quotation lists for the "La Coulisse" Stock Exchange in Paris during the 19th and 20th century.

We work on modeling strategies of analysis for sequential data for using the redundancy between the successive days of listing and the consistency between the global sequences of data.

Current activities are described in section 3.9.

5.3 Bilateral industry grants

5.3.1 Research contract Interdigital company

Participant: Christian Raymond, Bertrand Coüasnon, Yann Ricquebourg.

- Partners: *Interdigital company*
- Since 2020
- Contract: INSA

Intuidoc team started to work with Interdigital company on latent representations in deep learning. This collaboration is based on the CIFRE grant for the PhD of Martin Dornier.

Current activities are described in section 3.14.

5.3.2 Research contract Doptim company

Participant: Aurélie Lemaitre, Bertrand Coüasnon.

- Partners: *Doptim company*
- Since 2019
- Contract: INSA

Intuidoc team started to work with Doptim company in the field of document recognition applied to parish registers. This collaboration is based on a CIFRE grant.

Current activities are described in section 3.12.

5.3.3 Research contract Learn&Go company : KALUX (Kaligo Luxembourg)

Participant: Eric Anquetil, Nathalie Girard, Omar Krichen, Pauline Nerdeux.

- Partner: *Company Learn&Go*
- 12 months (2021-2022).
- Contract: INSA

After having developed the handwriting analysis engines for the French and English languages for the company Learn&Go within the scope of the Labcom Script&Labs (see section 5.2.4), we are now developing the analysis engines for numbers, upper and lower case letters for the German language in order to allow the company Learn&Go to release their Kaligo application in Luxembourg.

5.3.4 Industrial software licensing and technology transfer with Learn&GO company

Participant: Eric Anquetil, Nathalie Girard, Omar Krichen, Pauline Nerdeux.

- Partners: *Learn&GO company*
- 2017-2022
- Contract: INSA

The IntuiDoc team has close links with the Learn&GO company for transferring its research results for e-education. This partnership is now based on several licensing agreements. They cover various technologies of the Intuidoc team such as handwriting recognition and analysis.

This partnership is also supported by several collaborative projects: in particular the ANR LabCom "Script&Labs", the *IntuiScript* BPI Project (http://intuiscript.com/) and the P2IA project. This collaboration is also based on a CIFRE grant (A. Lods PhD Student) in relation with the ANR LabCom "Script&Labs".

The developments resulting from Labcom and P2IA have enabled the KALIGO product to be marketed in France, England and soon in Luxembourg and Germany, with high added value functionalities. The SATT OUEST VALORISATION is also supporting this technology transfer by participating in the equity of the company Learn&Go in 2021. The effective work of the INTUIDOC team has been fundamental to the technical evolution of the KALIGO product.

Current activities are described in sections 3.6, 3.5, and 3.1.

5.4 Collaborations

5.4.1 Rochester Institute of Technology, USA

Participants: Kwon-Young Choi, Bertrand Coüasnon, Yann Ricquebourg, Richard Zanibbi (RIT).

- Partner: Rochester Institute of Technology (Rochester, NY, USA).
- Since 2016

In collaboration with Richard Zanibbi from the Rochester Institute of Technology (RIT), Rochester, NY, USA, we work on interactive learning by combining deep learning technics, syntactical analysis and user interaction to introduce learning of segmentation. We propose to study the strong combination of deep learning and syntactical methods to build a document structure recognition system able to deal with segmentation problems by learning them. The syntactical part models the structure and brings complex context to the deep learning recognition. This collaboration is done through the co-supervising (Bertrand Coüasnon, Yann Ricquebourg and Richard Zanibbi) of the PhD of Kwon-Young Choi. PhD Defended on June 30th 2021 [1]

6 Dissemination

6.1 Promoting scientific activities

6.1.1 Scientific Events Selection

Member of Conference Program Committees

- E. Anquetil is member of the program committee of the 14th IAPR International Workshop on Graphics Recognition, GREC 2021
- E. Anquetil is member of the program committee of the International Conference on Document Analysis and Recognition (ICDAR 2021)

- B. Coüasnon is Senior member of the program committee of the International Conference on Document Analysis and Recognition (ICDAR 2021)
- B. Coüasnon is member of the program committee of the 14th IAPR International Workshop on Graphics Recognition, GREC 2021
- B. Coüasnon is member of the program committee of the 6th International Workshop on Historical Document Imaging and Processing (HIP 2021).
- B. Coüasnon is member of the program committee of the 3rd International Workshop on Open Services and Tools for Document Analysis (ICDAR-OST 2021).
- B. Coüasnon is member of the program committee of the Workshop on Industrial Applications of Document Analysis and Recognition (WIADAR 2021).
- B. Coüasnon is mentor of the Doctoral Consortium of the International Conference on Document Analysis and Recognition (ICDAR-DC 2021).
- A. Lemaitre is member of the program committee of the International Conference on Document Analysis and Recognition (ICDAR 2021).
- A. Lemaitre is member of the program committee of the 6th International Workshop on Historical Document Imaging and Processing (HIP 2021).

Reviewer

- Y. Soullard, K. Barrere, S. Corbillé and O. Krichen were reviewers of the International Conference on Document Analysis and Recognition (ICDAR 2021).
- S. Corbillé was reviewer of the British Machine Vision Conference (BMVC 2021).

6.1.2 Award

- K. Barrere received the Best Poster Award at the Doctoral Consortium of the International Conference on Document Analysis and Recognition (ICDAR-DC 2021) [3].

6.1.3 Scientific Expertise

- E. Anquetil was a reviewer in 2021 of one project for the French National Research Technology Agency (ANRT).
- A. Lemaitre was a reviewer in 2021 of one project for the French National Research Agency (ANR).

6.1.4 Research Administration

- E. Anquetil is a member of the executive committee of the society GRCE : " Groupe de Recherche en Communication Écrite ".
- Intuidoc members are members of the AFRIF (Association Française pour la Reconnaissance et l'Interprétation des Formes) and IAPR (International Association for Pattern Recognition) associations.
- E. Anquetil, B. Coüasnon, J. Camillerapp, A. Lemaitre and Y. Soullard, take part in the activities of the society GRCE : " Groupe de Recherche en Communication Écrite ".
- E. Anquetil is a member of the steering committee of LOUSTIC laboratory of Rennes (laboratoire d'observation des usages des technologies de l'information et de la communication).
- E. Anquetil is a member of the educational committee of the "DIGISPORT" University Research School (EUR).
- E. Anquetil is project manager for "Innovation and Entrepreneurship" at INSA Rennes. He is in charge of the construction of the student incubator project in Rennes "Station Rennes Innovation" which brings together 11 higher education establishments in Rennes.
- E. Anquetil is an elected member of the administration council of INSA Rennes.
- E. Anquetil is a member of the administration council of INSA Group Foundation.
- B. Coüasnon is member of the board of Valconum (Centre Européen de Valorisation Numérique).
- B. Coüasnon is an elected member of the laboratory council of the INSA component of IRISA.
- B. Coüasnon was scientific head of the Media and Interactions Department of IRISA, until October 1st 2021.
- B. Coüasnon was member of the scientific board of IRISA, until October 1st 2021.
- N. Girard is an elected member of the administration council of UFR ISTIC, Univ. Rennes 1.

6.2 Teaching, supervision

6.2.1 Teaching

The team is mainly made up of teachers who are very implied in activities of teaching. But a majority of lectures are not attached to this research topic, so they are not mentioned here.

- E. Anquetil is program manager of the Master of Science *"Innovation and Entrepreunership"* of INSA and Rennes School of Business (RSB).
- E. Anquetil and N. Girard give lectures at *Research in Computer Science (SIF)* Master of University of Rennes 1, University of Southern Brittany, ENS Rennes, INSA Rennes and CentraleSupélec.
- E. Anquetil is in charge of the module "Analysis, Interpretation and Recognition of 2D (touch) and 3D Gestures for New Man-Machine Interactions" (AIR) of the *Research in Computer Science (SIF)* Master of University of Rennes 1, University of Southern Brittany, ENS Rennes, INSA Rennes and CentraleSupélec.
- E. Anquetil is in charge of the module "Motion Analysis and Gesture Recognition (2D / 3D)" (AMRG) of the Computer Science Dept. of INSA Rennes.
- B. Coüasnon is co-Head with A. Termier of the *Research in Computer Science (SIF)* Master of University of Rennes 1, University of Southern Brittany, ENS Rennes, INSA Rennes and CentraleSupélec (https://master.irisa.fr).
- B. Coüasnon was in charge of the module on professionalization adapted to research (PROF) of the *Research in Computer Science (SIF)* Master of University of Rennes 1, University of Southern Brittany, ENS Rennes, INSA Rennes and CentraleSupélec, until June 2021.
- B. Coüasnon was invited for two courses at Master-Research "New technologies applied to History" of the Ecole nationale des Chartes on : "Digital Documents: Textual Documents" and "Automatic Access to Old Documents", Paris, France.
- Y. Soullard is in charge of the part "Text Mining and Deep Learning" of the module "Introduction to the Text Mining" at Master MAS *(Mathématiques Appliquées, Statistique (Science des Données))* of Rennes 2 University.

6.2.2 Supervision

- PhD in progress: T. Neitthoffer, Joint analysis of handwriting localization and recognition in structured documents, B. Coüasnon, A. Lemaitre, Y. Soullard, A. M. Awal (AriadNext), INSA Rennes, started May 2021.
- PhD in progress: K. Barrere, Deep neural networks and attention mechanisms for handwritten text recognition, B. Coüasnon, A. Lemaitre, Y. Soullard, INSA Rennes, started October 2020.
- PhD in progress: M. Dornier, Deep Learning in the Wild, B. Coüasnon, P.H. Gosselin, C. Raymond, Y. Ricquebourg, INSA Rennes, started October 2020.
- PhD in progress: F. Imbert, Design of a deep neural network architecture dedicated to the synthesis of handwriting from kinematic sensors of a sensors of a digital pen, E. Anquetil, Y. Soullard, R. Tavenard, INSA Rennes, started October 2021.
- PhD in progress: W. Mocaër, Spatio-Temporal Convolutional Neural Network for early action detection and analysis, E. Anquetil, R. Kulpa, INSA Rennes, started October 2020.

- PhD in progress: S. Corbillé, Hybridization of "Transparent" and "Deep Learning" AI approaches for automated handwriting analysis of children in the context of education, E. Anquetil, E. Fromont, Univ. Rennes 1, started October 2019.
- PhD in progress: S. Tarride, Combination of logical and textual knowledge for recognition of ancient register images, B. Coüasnon, A. Lemaitre, S. Tardivel (Doptim), INSA de Rennes, started February 2019.
- PhD in progress: C. Guerry, Historical big data: modelization of strategies to analyse collections of documents, B. Coüasnon, A. Lemaitre, S. Adam (Univ Rouen), INSA de Rennes, started October 2018.
- PhD in progress: C. Leroy, Incremental learning and evolving recognition system: application to on-the-fly recognition of handwritten gestures, E. Anquetil, N. Girard, INSA Rennes, started October 2018.
- PhD in progress: A. Lods, On-line analysis of handwritten arithmetic operation on digital tablet: Design of an innovative educational solution to improve learning arithmetic calculations in elementary school, E. Anquetil, S. Macé, INSA Rennes, started Febuary 2018.
- PhD Defended: K.Y. Choi, Combination of unsupervised generative models and a syntactical method for music symbol detection with few annotated data, B. Coüasnon, Y. Ricquebourg, R. Zanibbi (RIT, Rochester, USA), INSA de Rennes, started October 2016, defended June 30th 2021 [1].

6.2.3 Juries

- A. Lemaitre was reviewer in the thesis committee of Clément Sage PhD, Deep learning for information extraction from business documents, Université de Lyon, September 2021.

- B. Coüasnon was member of the thesis committee of Vincent Poulain d'Andecy's PhD, Système à connaissance incrémentale pour la compréhension de document et la détection de fraude, Université de La Rochelle, October 2021.

6.3 Patent and Deposit of digital creations (APP)

- E. Anquetil, D. Simmonet and M. Renault deposited a V3 of the digital creation, ISA - Isolated Symbol Analysis.(ISA, 19/02/2021)

- E. Anquetil, D. Simmonet and M. Renault deposited a V3 of the digital creation, IWA – Isolated Word Analysis.(IWA, 19/02/2021)

7 Bibliography

Doctoral dissertations and "Habilitation" theses

[1] K.-Y. Choi, *Combination of unsupervised generative models and a syntactical method for music symbol detection with few annotated data*, Theses, INSA de Rennes, June 2021.

Articles in referred journals and book chapters

[2] S. Tarride, A. Lemaitre, B. B. Coüasnon, S. Tardivel, "Combination of deep neural networks and logical rules for record segmentation in historical handwritten registers using few examples", *International Journal on Document Analysis and Recognition*, January 2021, https://hal.archives-ouvertes.fr/hal-03160212.

Publications in Conferences and Workshops

[3] K. Barrere, Y. Soullard, A. Lemaitre, B. Coüasnon, "Transformers for Historical Handwritten Text Recognition", *in : 16th International Conference on Document Analysis and Recognition (ICDAR 2021) Doctoral Consortium*, Nibal Nayef and Jean-Christophe Burie, Lausanne, Switzerland, September 2021, https://hal.archives-ouvertes.fr/hal-03485262.

[4] O. Krichen, S. Corbillé, E. Anquetil, N. Girard, P. Nerdeux, "Online analysis of children handwritten words in dictation context", *in : 14th International Workshop on Graphics Recognition*, Lausanne, Switzerland, September 2021, https://hal.archives-ouvertes.fr/hal-03448357.

[5] A. Lemaitre, J. Camillerapp, "Segmentation of historical maps without annotated data", *in : 6th International Workshop on Historical Document Imaging and Processing (HIP'21)*, *The 6th International Workshop on Historical Document Imaging and Processing*, Association for Computing Machinery, p. 19–24, Lausanne, France, September 2021, https://hal.inria.fr/hal-03374571.

[6] C. Leroy, E. Anquetil, N. Girard, "Drift anticipation with forgetting to improve evolving fuzzy system", *in : 25th International Conference on Pattern Recognition (ICPR2020)*, Milan, Italy, January 2021. https://arxiv.org/abs/2101.02442, https://hal.archives-ouvertes.fr/hal-02974253.

[7] A. Lods, E. Anquetil, S. Macé, "Segmentation and graph matching for online analysis of student arithmetic operations", *in : 16th International Conference on Document Analysis and Recognition*, Lausanne, Switzerland, September 2021, https://hal.archives-ouvertes.fr/hal-03259438.

[8] W. Mocaër, E. Anquetil, R. Kulpa, "Online Spatio-Temporal 3D Convolutional Neural Network for Early Recognition of Handwritten Gestures", *in : ICDAR 2021 - 16th International Conference on Document Analysis and Recognition*, p. 1–16, Lausanne, Switzerland, September 2021, https://hal.archives-ouvertes.fr/hal-03229957.