Symposium International Francophone sur l'Ecrit et le Document

SIFED 2023

Résumés des contributions

Journée du 9 juin 2023
Organisée à Paris, Université Paris Cité (France)

Table des matières

1 Jeu de données de tickets de caisse pour la détection de fraude documentaire

Mots-clefs : Fraude documentaire, jeu de données, détection de fraude

Beatriz Martínez Tornés, Théo Taburet, Emanuela Boros, Kais Rouis, Petra Gomez-Krämer, Nicolas Sidere, Antoine Doucet and Vincent Poulain D'Andecy

Résumé : L'utilisation généralisée de documents numériques non sécurisés par les entreprises et les administrations comme pièces justificatives les rend vulnérables à la falsification. En outre, les logiciels de retouche d'images et les possibilités qu'ils offrent compliquent les tâches de la détection de fraude d'images numériques. Néanmoins, la recherche dans ce domaine se heurte au manque de données réalistes accessibles au public. Dans cet article, nous proposons un nouveau jeu de données pour la détection des faux tickets contenant 988 images numérisées de tickets et leurs transcriptions, provenant du jeu de données SROIE (scanned receipts OCR and information extraction). 163 images et leurs transcriptions ont subi des modifications frauduleuses réalistes et ont été annotées. Nous décrivons en détail le jeu de données, les falsifications et leurs annotations et fournissons deux baselines (basées sur l'image et le texte) sur la tâche de détection de la fraude.

2 Tables de recensement historiques : un défi de segmentation de document et de reconnaissance d'écriture manuscrit

Mots-clefs : Documents historiques Humanités numériques Document Image Understanding

Guillaume Bernard and Casey Wall

Résumé : **Contexte** Les recherches scientifiques en démographie s'appuient en particulier sur des corpus de documents pour analyser par exemple les mouvements de population à travers un territoire sur un période définie. Les tables de recensement historiques renseignent les naissances, mariages ou décès enregistrés au sein des localités. Traitées en lot par des moyens numériques, ces données permettraient d'identifier, entre autres, de grands mouvement de population de contextualiser des migrations ou l'apparition de patronymes, par exemple. Pour ce projet de recherche en informatique, nous cherchons à résoudre les verrous scientifiques de ce type de document dont l'extraction du contenu manuscrit de ces tables, parfois dégradées. **Tables de recensement historiques** Nous disposons d'un corpus de plus de 537 numérisations avec transcriptions de tables de recensement historiques provenant de deux communes de France : Echevronne (Côté d'Or) et Vic-sur-Seille (Moselle). Ces documents proviennent des archives nationales. Ils présentent l'ensemble des dégradations remarquables : les lignes sont majoritairement rayées et annotées au crayon à papier, certains termes sont raturés et remplacés par d'autres pour corriger une erreur. Le formalisme du modèle de la page n'est pas toujours respecté et des informations devant être renseignées dans une colonne le sont dans une autre. Enfin, le support d'origine peut avoir été fortement dégradé. Du ruban adhésif maintient les bords extérieurs de la page, la teinte de la colle est visible, etc. **Défis** scientifiques Ces tables historiques numérisées concentrent différents type de problèmes. En tant que documents historiques numérisés, la dégradation du support et la qualité des numérisations influent en premier lieu. Pour tout document, le processus d'analyse consiste à identifier les sections pertinentes et les extraire de l'image : ce sont les lignes manuscrites. Chaque ligne correspond à un enregistrement dans la table de recensement. Une fois extraites, car reconnues, les lignes sont analysées par des outils de reconnaissance manuscrite pour en extraire le texte. En dehors des défis purement scientifiques que représente l'analyse de ces documents au niveau informatique, le volume des données à traiter est également, en lui-même, un enjeu de recherche. Le traitement des numérisations en volume impose une vitesse d'exécution proche du temps réel pour les traiter en lots dans des temps raisonnables et qu'ils soient exploités par les démographes, utilisateurs finaux des outils développés au sein de ce projet de recherche. **Conclusion** Ce projet répond à des problématiques concrètes de sciences humaines et sociales tout en intégrant des questions de recherche en informatique centrées sur l'analyse de document historique. Nous cherchons à faire avancer l'état de l'art sur ces deux éléments : détection de lignes dans les tables historiques et reconnaissance du texte manuscrit. Nous ajoutons la contrainte des temps d'inférence pour rendre notre approche utilisable dans le traitement des les volumes massifs de documents disponibles.

3 Grouping and linking Key/Value pairs in business documents using expert methods and pretrained transformers

Mots-clefs : Key/value extraction Business documents Pretrained transformers

Eliott Thomas

Résumé : Key/value extraction is a challenging task in document AI, especially in business documents like invoices. Accurately extracting key/value pairs from business documents is crucial for enabling downstream processing tasks such as accounting, analytics, and decision-making. We propose a method for grouping and linking key/value pairs in business documents using expert methods and pretrained transformers. One of the main challenges in key/value extraction is the diverse and com- plex layouts of business documents, which can make it difficult for automated methods to accurately identify and extract the relevant information. We demon- strate the effectiveness of our method by presenting results on two datasets : a private dataset from our company and a public dataset, XFUND. Our method comprises three steps : 1. Grouping words into groups of words based on expert methods that rely on layout information obtained via OCR. This step addresses the complexity and diversity of layouts in business documents, which can pose challenges for automated methods. 2. Classifying the groups of words into key, value, or other using a pretrained BERT multilingual model. This step enables us to handle documents in different languages and improves the precision of key/value extraction. 3. Linking the predicted key/value pairs based on their relative positions in the document, leveraging the layout information. Our approach achieves precision scores similar to those of state-of-the-art methods with precision scores of around 0.82 and 0.75 for classification and link- ing on the XFUND dataset, respectively. Our results demonstrate the effective- ness of combining expert methods with pretrained transformers for key/value extraction in business documents.

4 SET, SORT ! A Novel Sub-Stroke Level Transformers for Offline Handwriting to Online Conversion

Mots-clefs : offline handwriting transformer pen trajectory recovery

Elmokhtar Mohamed Moussa, Thibault Lelore and Harold Mouchère

5 VLCDoC : Vision-Language Contrastive Pre-Training Model for Cross-Modal Document Classification

Mots-clefs : Document Representation Learning Classification Contrastive Learning

Souhail Bakkali, Mickael Coustaty, Zuheng Ming, Oriol Ramos Terrades and Marçal Rusiñol

Résumé : Multimodal learning from document data has achieved great success lately as it allows to pre-train semantically meaningful features as a prior into a learnable downstream task. We approach the document classification problem by learning cross-modal representations through language and vision cues, considering intra-modality and inter-modality relationships. Instead of merging features from different modalities into a joint representation space, the proposed method exploits high-level interactions and learns relevant semantic information from effective attention flows within and across modalities. The proposed learning objective is devised between intra-modality and inter-modality alignment tasks, where the similarity distribution per task is computed by contracting positive sample pairs while simultaneously contrasting negative ones in the joint representation space. Extensive experiments on public document classification datasets demonstrate the effectiveness and the generality of our model on low-scale and large-scale datasets.

6 Entraînement d'architectures type transformer sur peu de données : application à la reconnaissance de texte manuscrit ancien

Mots-clefs : Architectures légères Documents anciens Transformer

Killian Barrere, Yann Soullard, Aurélie Lemaitre and Bertrand Coüasnon

Résumé : Les architectures type Transformer donnent d'excellents résultats pour la reconnaissance de textes manuscrits et sont devenues l'architecture standard pour reconnaître des documents modernes. Cependant, elles nécessitent des quantités importantes de données annotées pour obtenir des résultats compétitifs. Elles s'appuient généralement sur des données synthétiques pour résoudre ce problème. La reconnaissance de textes manuscrits anciens représente un défi en raison des dégradations, des écritures spécifiques pour lesquelles peu d'exemples sont disponibles et des langues anciennes qui varient au fil du temps. Ces limitations rendent également difficile la génération de données synthétiques réalistes. Avec des données suffisantes et appropriées, les architectures type Transformer pourraient atténuer ces problèmes, grâce à leur capacité à avoir une vue globale des images textuelles et à leurs capacités de modélisation du langage. Dans cet article, nous proposons l'utilisation d'un modèle de Transformer léger pour s'attaquer à la tâche de reconnaissance de textes manuscrits historiques. Pour entraîner l'architecture, nous introduisons des données synthétiques réalistes reproduisant le style des écritures historiques. Nous présentons une stratégie spécifique, à la fois pour l'entraînement et la prédiction, afin de traiter les documents historiques, pour lesquels seule une quantité limitée de données d'entraînement est disponible. Nous évaluons notre approche sur l'ensemble de données READ de l'ICFHR 2018 qui est dédié à la reconnaissance d'écriture manuscrite dans des documents anciens spécifiques. Les résultats montrent que notre approche type Transformer est capable de surpasser les méthodes existantes.

7 Adaptation de domaine pour la reconstruction de la trajectoire du stylo à partir de capteurs cinématiques

Mots-clefs : Adaptation de domaine reconstruction de trajectoire unité de mesure inertielle

Florent Imbert, Yann Soullard, Romain Tavenard and Eric Anquetil

Résumé : L'écriture manuscrite à l'aide d'un stylo connecté devient l'une des principales méthodes d'interaction entre l'homme et l'ordinateur. Par rapport aux systèmes traditionnels d'écriture manuscrite sur écran tactile, le stylo présente l'avantage de produire un signal d'écriture manuscrite en ligne, sans contraintes de surface. En effet, les personnes qui écrivent sur un papier obtiennent les coordonnées correspondantes à la trajectoire du stylo qui représentent le signal d'écriture en ligne. De plus, la sensation d'écrire sur du papier est importante, en particulier pour les enfants pendant l'apprentissage de l'écriture. Dans ce travail qui fait partie du projet ANR franco-allemand KIHT avec Stabilo, nous introduisons une approche basée sur l'adaptation de domaine qui reconstruit l'écriture sur papier du stylet numérique Digipen de STABILO. Ce dernier est équipé d'un système de suivi de trajectoire sans fil basé sur des capteurs cinématiques. Nous utilisons une méthode d'adaptation de domaine non supervisée, pour passer du domaine de la tablette où la vérité terrain (la trace en ligne de l'écriture sur la tablette) est connue, au domaine du papier où seules les données des capteurs d'entrée sont connues.

8 Intégration de Connaissance au sein d'un Réseau Multi-tâches pour l'Analyse de Nouveaux Types de Documents d'Identité

Mots-clefs : Knowledge Integration Multitask Learning OCR Text Localization Identity Documents

Timothée Neitthoffer, Aurélie Lemaitre, Bertrand Couasnon, Yann Soullard and Ahmad Montaser Awal

Résumé : La reconnaissance de Documents d'Identité est une étape clé pour les applications de Know Your Customer (KYC) où les Documents d'Identité sont vérifiés. Les documents appartenant au même type de document (carte d'identité, passeport ou autre, issu du même pays et appartenant à la même version) partagent la même structure de champs, que nous appelons modèles. Les systèmes traditionnels de reconnaissance de documents utilisent ce modèle afin de guider la localisation des champs et ensuite leur reconnaissance. Cependant, ces systèmes doivent être ajustés aux différents modèles afin de correctement les traiter. Ainsi, ces systèmes ne peuvent être utilisés directement pour de nouveaux modèles de documents. Dans ce travail, nous traitons la tâche de localisation et reconnaissance du texte dans le contexte de nouveaux documents, où seul le modèle est disponible, sans exemples labelisés pour l'apprentissage. Pour cela, nous proposons de s'appuyer sur le modèle comme une entrée supplémentaire du réseau afin de conditionner ses prédictions. Ainsi, celui-ci apprends à se baser sur le modèle pour ne pas avoir besoin de données labélisées pour s'adapter aux nouveaux modèles. Nous définissons alors trois concepts importants pour l'intégration de connaissances au sein des réseaux de neurones et les explorons dans notre cadre des documents d'identités. Nous adaptons également un réseau multitâche afin d'effectuer de manière jointe la localisation et la reconnaissance du texte. Afin d'évaluer notre approche, nous concevons un nouveau jeu de données et une nouvelle tâche pour la base de données publique MIDV2020, à partir de photos "in-the-wild" rectifiées. Notre méthode obtient les meilleurs résultats sur cette nouvelle tâche pour deux jeux de données dont un industriel composé d'exemples réels.

9 Improving Information Extraction from Semi-Structured Documents Using Attention based Semi-variational Graph Auto-encoder

Mots-clefs : Semi-Structured Document Multi-GAT VGAE Labeled and Unlabeled document

Djedjiga Belhadj, Abdel Belaïd and Yolande Belaïd

Résumé : In this work, we propose a semi-supervised system for information extraction from administrative documents, that learns from both labeled and unlabeled data. The document is modeled as a words graph, where each node contains the textual, layout and visual features of the word and it is connected to its spatially close neighbors. Semi-supervised variational graph auto-encoders (VGAE) have proven efficient on graph-based tasks, but they usually separate the classifier from the encoder and decoder and don't take full advantage of the VGAE model for the benefit of the classification. To optimize the classification as much as possible, we propose a semi-VGAE with an attention-based classifier that shares its layers with the VGAE encoder. This is further enhanced by proposing a VGAE loss managed by the classification loss. Experiments show that our model helps improve nodes prediction accuracy. We tested the architecture on two artificially generated datasets : Gen-Invoices and Gen-Payslips and one real dataset : receipts issued from the SROIE ICDAR 2019 competition. The latter dataset yielded an important F1 score of 97.94

10 Sécurité d'un document par un hologramme

Mots-clefs : Hologram Identity Documents Image and Video analysis Authentication

Camille Kurtz and Nicole Vincent

Résumé : Pour assurer la sécurité des documents comme un passeport ou un billet de banque, des hologrammes sont introduits dans les documents. En effet, ils sont difficilement reproductibles avec une imprimante classique. Actuellement la vérification de la présence d'un hologramme est réalisée à l'œil nu, en lumière naturelle ou avec des lumières particulières. Avec l'évolution des modes d'échanges de documents qui sont de plus en plus souvent réalisés de manière électronique, il devient indispensable d'analyser de manière automatique les images d'hologrammes. L'objectif de l'étude est de détecter la présence d'un hologramme dans une vidéo acquise à l'aide d'une caméra embarquée sur un téléphone portable, permettant rapidement de détecter un faux document d'identité. Nous présentons ici une méthode ne s'intéressant pas aux trames individuellement mais considérant globalement la suite des trames au cours de la vidéo. Le système proposé repose sur un apprentissage de bout en bout, il est constitué d'un réseau neuronal profond calculant des caractéristiques spatio-temporelles et un classifieur permettant la prise de décision. La méthode détecte, sur l'ensemble de la vidéo, des incohérences liées à la présence d'un hologramme potentiel. À partir de cette étude spatio-temporelle, un retour à l'aspect spatial permet de détecter les pixels ou les zones correspondant à une présence d'hologramme. Cette étape peut être considérée comme une étape intermédiaire avant l'interprétation de l'hologramme, la reconnaissance du pays d'origine du passeport par exemple. Les résultats seront donnés sur la base de vidéos d'hologrammes acquises à l'aide de l'appareil photo d'un téléphone portable, mais aussi sur une base de vidéos accessible à la communauté (comme MIDV-500).

11 Identity document layer decomposition

Mots-clefs : semantic segmentation deep learning identity document layer decomposition

Glen Pouliquen, Guillaume Chiron and Montaser Awal

Résumé : This study takes place in an industrial context and aims at improving services for remote verification of identity documents. This work is exploratory and involves photos/videos of identity documents mainly captured by smartphone. Being able to extract/decompose layers from these ID document images has many advantages : a better decoding of textual fields by running the OCR on a "text only" layer ; the ability to generate fake documents by merging generated and real layers ; the ability to reduce noise or to anonymise a document by removing/fitlering some layers (ex : text and photo). The ultimate goal is the development of a single model capable of extracting these layers, each containing the signal of a specific element, which could be categorized in 4 types : a) structural (ex : text fields, photos, signatures) ; b) visually variable security elements (ex : OVD, hologram, OVI) ; c) noise (ex : reflects, flash) ; d) background (without any previously listed elements). Firstly, we established a train/evaluation dataset along with required annotations. Then, a deep learning model for semantic segmentation (pixel level classification) is trained to obtain the masks representing different layers. Finally, the different layers are generated, either by combining the previous masks with inpainting methods, or directly by training generative models for the image-to-image translation. In practice, the contribution focused mainly on the second stage, proposing models for segmenting the structural elements but also noise and variable security elements. A pipeline has also been proposed to generate layers for the security elements, the variable text (name, surname etc), the pictures, the signature, and the background of the identity document.

12 Node Induced Sub-graph Method for Information Extraction from Administrative Documents

Mots-clefs : Graph BERT Information Extraction

Dipendra Sharma Kafle, Aurélie Joseph and Mickael Coustaty

Résumé : Information Extraction (IE) plays a pivotal role in the automation of auditing process in business documents. The goal is to find and extract words related to important data such as dates and amounts. However, variety in text layouts of documents is a challenge to extract the information with utmost accuracy. In this research, we use Graph method to properly map the data in the document and extract tags from invoice like documents. We use attention mechanism in graph and consider this IE task as node classification in graphs, where nodes are words in the document. We train on multiple node-induced sub-graphs for more enriched knowledge and generalized learning. In comparision, our approach can improve prediction results for some of the tags when compared to related works with BERT, graphs and general classification methods.

13 Système tutoriel intelligent pour l'apprentissage par croquis (SKETCH)

Mots-clefs : Systèmes tutoriels intelligents Interprétation de documents manuscrits semi-structurés

Islam Barchouch, Omar Krichen, Eric Anquetil and Nathalie Girard

Résumé : Ce résumé présente les premiers travaux réalisés dans le cadre du projet ANR SKETCH. Ce projet vise à développer un système tutoriel intelligent d'apprentissage par croquis, pour les étudiants dans les formations de Santé. L'objectif est d'accompagner les étudiants dans la réalisation de schémas d'anatomie à main levée. Pour cela on propose un environnement d'apprentissage interactif et dynamique qui repose sur l'utilisation de tablettes avec stylet. L'idée consiste à simuler l'approche traditionnelle papier/crayon pour faciliter le transfert de l'apprentissage du support numérique vers le support papier. Le système offrira d'une part la création d'exercices par les enseignants via un mode auteur et d'autre part, l'interprétation et le suivi dynamique des réalisations via un mode étudiant. Nos premiers travaux se concentrent sur cette seconde partie d'interprétation et d'analyse des croquis des étudiants. Ces croquis manuscrits sont semi-structurés et se caractérisent par un large éventail de styles de production. Selon le domaine d'anatomie adressé ou la typologie de l'exercice, on pourra ainsi s'intéresser soit à évaluer précisément la qualité des formes réalisées, ou vérifier la cohérence sémantique et procédurale du schéma. Pour relever ce défi, le système s'appuie sur des techniques de reconnaissance de formes pilotées par des règles de composition grammaticale bidimensionnelles permettant de modéliser la structure du document. Pour s'adapter à la liberté de dessin, tout en gardant une mesure sur la validité des croquis, la reconnaissance des formes reposera sur des IA de reconnaissance incrémentale et évolutive. La stratégie consiste à développer, à partir de cette reconnaissance, une analyse des compositions manuscrites de l'étudiant au regard d'un modèle de référence établi par l'enseignant. Ce modèle sera défini par un graphe de connaissances extrait à partir du dessin de l'enseignant et qui modélise le savoir du domaine. L'objectif est de repérer les erreurs de réalisation ou les imprécisions commises par l'étudiant, ou de l'aider en lui fournissant le guidage nécessaire s'il en a besoin. À cette fin, différents types de feedback seront proposés à l'étudiant. À terme, l'objectif sera de déduire de manière dynamique les règles de composition à partir des schémas de référence fournis par l'enseignant dans le mode auteur. Nous présenterons également les résultats préliminaires des expérimentations effectuées avec nos collègues chercheurs en psychologie expérimentale (laboratoire LP3C, Univ Rennes 2) auprès des étudiants de l'IFPEK* sur la 1ère version du système.

14 Correction post-OCR en trois étapes : détection, correction, validation

Mots-clefs : transcription post-OCR correction NLP seq2seq

Guillaume Thomas, Joseph Chazalon and Edwin Carlinet

Résumé : La correction post-OCR est une solution de dernier recours utilisée lorsque la spécialisation d'un système de transcription n'est pas réalisable : données images absentes, entraînement trop complexe... Elle consiste à retrouver la chaîne de texte correcte la plus probable à partir de résultat d'un système de transcription (OCR ou HTR) contenant des erreurs. Face à la quantité limité de solutions ouvertes permettant d'attaquer facilement ce problème, nous proposons une implémentation libre et open-source d'un système de correction post-transcription en trois étapes : détection d'erreurs, correction d'erreurs et contrôle des corrections apportées. Ce système est simple à entraîner et utiliser, et ne requiert de connaissance en Python que superficielles. Ce système a permis de réduire le nombre d'erreurs au niveau caractère sur un jeu de données d'annuaires du commerce du 19e siècle, en Français, déjà très bien reconnu et comportant des nombres. Nous profiterons de cette présentation pour proposer une analyse des forces et faiblesses de ce type d'approche, et présenter des résultats sur d'autres jeux de données.

15 Deep learning appliqué au traitement des ordonnances médicales

Mots-clefs :

Jonhatan Pattin-Cottet, Véronique Eglin, Alexandre Aussem

Résumé :

16 Extraction de données sensorielles dans des documents hétérogènes texte-image

Mots-clefs :

Cédric Boscher, Véronique Eglin

Résumé : Notre travail vise à proposer une approche permettant d'étendre la recherche d'information multimodale à cinq sensorialités (la vue, le toucher, l'odorat, l'ouïe, le goût et l'odorat), applicable à des collections de documents anciens, dans un contexte multimodal texte et image. Afin de pouvoir interroger des sources de documents texte et images avec des requêtes texte, nous proposons d'améliorer des tâches de recherche d'information multimodale telles que l'Image-Text Matching qui consiste à aligner une requête texte avec l'image la plus proche sémantiquement dans une collection finie d'images, ou encore les tâches de Visual Grounding References Expressions consistant à cibler la région d'intérêt d'une image la plus proche sémantiquement d'une requête. Nous nous appuyerons sur des architectures de modèles multimodaux pré-entrainés multi-modaux dits Vision-and-Language, ainsi que sur une modalité dérivée du texte, dite sensorimotrice, modélisant sémantiquement les concepts à travers le prisme de la multi-sensorialité. Afin de proposer une approche extensive multisensorielle, nous envisageons de définir et appliquer des tâches d'apprentissage par transfert cross-sensorielles basées sur des mécanismes d'attention, permettant d'utiliser un corpus conséquent dans une sensorialité source (l'odorat) comme base d'apprentissage permettant d'améliorer la classification de références vers d'autres sensorialités, pour lesquelles nous ne disposons pas de jeux de données conséquents à ce jour.

17 Amélioration de la reconnaissance d'entités textuelles dans les documents techniques par clustering incrémental

Mots-clefs :

Mathieu François, Véronique Eglin

Résumé : The digital transformation of engineering documents is an ambitious research topic in the industrial world. The representation of component identifiers (tags), which are textual entities without a language model is one of the major challenges. Most of OCR use dictionary-based correction methods so they fail at recognizing hybrid entities composed by numerical and textual characters. This study aims to adapt OCR results on language-free strings with a specific semantics and requiring an efficient post-OCR correction with unsupervised approaches. We propose a two-step methodology to face the questions of post-OCR correction in engineering documents. The first step focuses on the alignment of OCR transcriptions producing a single prediction refined from all OCR predictions. The second step presents a combined incremental clustering & correction approach achieving a continuous correction of tags' transcriptions relatively to their assigned cluster. For both steps,the dataset is produced from a set of 1600 real technical documents and placed at the disposal of for the research community. When compared to the best state-of-art OCR, the post-OCR approach produced a gain of 9