INSTITUT NATIONAL DES SCIENCES APPLIQUÉES RENNES

Demande de Contrat Doctoral Ordinaire - INSA RENNES 2024

FICHE DESCRIPTIVE (3 pages maximum*) – annexes autorisées (pdf)

Catégorie demande	X contrat doctoral 100 %
Si demande effectuée sur l'appel Contrats Doctoraux Ordinaires (ED SPI.bzh, S3M, Matisse) indiquez-le	☐contrat doctoral 100 % ☐contrat doctoral 50 %
Si cofinancement nature du cofinancement**	**(partenaire/organisme, montant cofinancement, nature du contrat, durée, calendrier)**
Titre du sujet de thèse	Generic Optical Music Recognition Combining complex syntactic rules and end-to-end systems
Composante/équipe	☐Inst. FOTON ☐ISCR-CSM ☐ISCR-CM ☐ISCR-COrint ☐IETR XIRISA ☐IRMAR ☐LGCGM ☐MECAPROCE
Direction de la thèse/Codirection (HDR)	Bertrand Coüasnon, Maitre de conférences HDR, INSA Aurélie Lemaitre, Professeur, Université Rennes 2
Co-encadrement	Denis Coquenet, Maitre de conférences, Univ Rennes
Partenaires impliqués	Aucun
Cotutelle / co-direction internationale envisagée* ?	Aucun
Description des objectifs scientifiques du projet de thèse	**Context:** The Shadoc team of the IRISA lab (http://www.irisa.fr/intuidoc) works on Hybrid Systems for Document Recognition. We mainly focus on the recognition of structure and content of ancient, damaged and handwritten documents (archive registers, newspapers, mails, musical scores…) This thesis focuses on the specific application of Optical Music Recognition (OMR. The goal is to take as input an image of a score, and to produce as output a digital version of this score, such as Midi or MusicXML file. Commercial OMR software is available. However, it produces unsatisfactory results. First, it must be applied to clean, recent documents, formatted in a standard style. Second, it still produces a large number of recognition errors, requiring a tedious manual correction stage. When a user wants to process a complex old partition, it is often simpler to input everything by hand than to use an OMR, which would produce a large number of errors. **Existing approach:** The Shadoc team has been working for a long time on music recognition. More recently, in the context of ANR Collabscore project, we have set up a system that combines deep learning and syntactic rules for music recognition. On the one hand, deep-learning approaches are convenient to detect symbols that are always similar: note heads, keys, accidentals, rests… That is why we have trained a system to recognize those elements. The difficulty is the need to adapt to new kinds of documents with few available data. On the other hand, the musical content can be described by precise rules: a score contains staff systems and bars with notes on them. The number of beats in a bar follows precise rules, depending on the notes used: round, half, quarter, eighth, etc. All these syntactic elements are described using the grammatical DMOS method, developed in the team. The great interest of using syntactic rules is the ability to detect inconsistencies in the produced results, and to solicit the user on precise errors of recognition.

	However, this method still needs to be improved to provide a generic OMR. **Scientific objective:** In the domain of document recognition, the recent work shows the interest of end-to-end approaches: those systems based on Transformer models, enable to recognize the text in images at a full-page level, without the need of the segmentation step. These approaches obtain very good results when training is performed with a large amount of annotated data, but the efficiency of such approach remains to be studied for databases with a limited amount of real training examples. Some researchers started to apply end-to-end methods to musical scores, but their work is limited to monophonic or piano scores. End-to-end deep learning systems are really interesting, but we believe that the knowledge on music structure is really important for understanding, and it deserves to be included in the recognition system. The questions raised in this thesis are: how to combine syntactical rules, such as musical score description, with an end-to-end approach? Is it necessary to follow a sequential approach, or how can knowledge of musical syntax be introduced at the input of an end-to-end system? How to make end-to-end approaches competitive with limited annotated data? **Application:** The goal is to provide a generic OMR that can be applied on: both recent and historical scores, monophonic to orchestra scores, printed and handwritten documents. The work will be applied to public research databases, as well as on the databases of Collabscore project, provided by the BNF (French national library).
Références	Jorge Calvo-Zaragoza, Alejandro H. Toselli, Enrique Vidal. Handwritten Music Recognition for Mensural notation with convolutional recurrent neural networks, Pattern Recognition Letters, Volume 128, 2019, Pages 115-121,https://doi.org/10.1016/j.patrec.2019.08.021. Denis Coquenet, Clément Chatelain, Thierry Paquet, "DAN: a Segmentation-free Document Attention Network for Handwritten Document Recognition", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023. A. Ríos-Vila, D. Rizo, J.M. Iñesta *et al.* "End-to-end optical music recognition for pianoform sheet music". *IJDAR* **26**, 347–362 (2023). https://doi.org/10.1007/s10032-023-00432-z Ali Yesilkanat, Yann Soullard, Bertrand Coüasnon, Nathalie Girard. Full-page music symbols recognition: state-of-the-art deep models comparison for handwritten and printed music scores. 2023. ⟨hal-04268139⟩ Kaiyang Zhou, Jingkang Yang, Chen Change Loy and Ziwei Liu, "Learning to Prompt for Vision-Language Models", *International Journal on Computer Vision*, 2022.
Positionnement thématique scientifique INSA	**☐ Matériaux, composants et procédés pour une énergie et une production industrielle soutenables /durables** **☐ Technologies pour la santé, le bien-être et qualité de vie** **X Numérique et communications sécurisés et responsables**