

Deep Learning Optical Music Recognition Systems

Keywords:

AI, Deep learning, Music recognition, End-to-end systems, Transformers

Context:

The Shadoc team at IRISA lab specializes in hybrid systems for document recognition, with a particular focus on ancient, damaged, and handwritten documents. This internship focuses on the Optical Music Recognition (OMR) task, *i.e.*, the transcription of digitized music scores into a numerical format.

Nowadays, two main trends exist: rule-based and end-to-end strategies.

In ruled-based approaches, the analysis is guided by syntactic and logical rules, that follow the music syntax. Thus, musical symbols are extracted in documents, and organised following a consistent organisation. This kind of approach was successfully applied in the previous project Collabscore [2], with the DMOS method [4].

The end-to-end approaches emerged as an interesting alternative, performing the recognition on a single autoregressive pass through an iterative implicit segmentation.

Initially proposed for text recognition [1], it was recently transposed to OMR with success [3]. The main advantage of such systems is that they do not require manual rule writing and can potentially offer better generalization and adaptability. However, they generally necessitate more training data, relying on synthetic document generation.

In addition to the reconnaissance strategy to be employed, there is also the question of the format of the ground truth. There is currently no consensus on the appropriate output format to use, as this may vary depending on the ultimate goal.

Thus, the first goal of this internship is to study the state-of-the-art approaches under several perspectives, to answer the following questions:

- Can end-to-end approach be used with a very limited amount of annotated data?
- In which cases traditional ruled-based approaches should be preferred to end-to-end systems?
- What are the limits of the Kern output format, compared to other formats such as MusicXML?

Depending on the internship progress and results, the submission of a scientific article to an international conference may be considered. Thesis funding may also be considered at the end of this research internship.

Références

[1] Denis Coquenot, Clément Chatelain, Thierry Paquet, "DAN: a Segmentation-free Document Attention Network for Handwritten Document Recognition", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023.

[2] Philippe Rigaux, Bertrand Coüasnon, Christophe Guillotel-Nothmann, Fabien Guilloux, Aurélie Lemaitre, The CollabScore Dataset -Towards Robust and Generalized OMR Evaluation, *13th International Conference on Digital Libraries for Musicology (DLfM 2026)*, Jul 2026, Thessalonique, Greece.

[3] Ríos-Vila A., Calvo-Zaragoza J., Rizo D., Paquet T. End-to-End Full-Page Optical Music Recognition for Pianoform Sheet Music. *Int J Comput Vis* **134**, 49 (2026).
<https://doi.org/10.1007/s11263-025-02654-6>

[4] Aurélie Lemaitre, Bertrand Coüasnon, Philippe Rigaux. CollabScore OMR: A Hybrid System for Music Score Recognition. *DAS 2026 - International Workshop on Document Analysis Systems*, Sep 2026, Vienne, Austria. pp.146 - 162, [10.1007/978-3-032-36207-0_9](https://doi.org/10.1007/978-3-032-36207-0_9)

Supervision:

Aurélie LEMAITRE, IRISA/Shadoc, Université Rennes 2, aurelie.lemaitre@irisa.fr
Denis COQUENET, IRISA/Shadoc, Université de Rennes, denis.coquenot@irisa.fr
Bertrand COUASNON, IRISA/Shadoc, Université de Rennes, bertrand.couasnon@irisa.fr

Location: IRISA – Rennes