ODE TRANSFORMERS FOR DOCUMENT INTELLIGENCE IN LOW-RESOURCE ENVIRONMENTS

Advisors: Oriol Ramos (oriolrt@cvc.uab.cat), David Romero (dromero@crm.cat)

Abstract
Digital transformation of public administration, industry and small and medium enterprises involve the development of efficient and sustainable document management processes. Automatizing document management process for information extraction can reduce human efforts in repetitive and human-error prone tasks, saving time, operational and energetic costs. In this context, the ability to automatically read, understand and interpret documents, Document Intelligence (DI), is a challenging task.

In the document analysis group, we do research on document intelligence systems where we integrate contextual knowledge with the hypothesis that the systems will improve their ability to understand and process a wide variety of document types. Within this line of research, we are interested in addressing the problem of the limitations of current models in a use case where the number of documents is small and/or their structural variability too complex. Specifically, we are interested in studying how document intelligence systems can be adapted to manage changing or evolving contexts, ensuring solid performance in out-of-distribution (OOD) data and what techniques can be used to optimize computational resources in document processing, making systems scalable and efficient in low-resource environments.

Inspired by current trends in Natural Language Processing (NLP), state-of-the-art in DI is dominated using pre-trained language models, based on the transformer architecture. Precisely, following [1,2], one can see that transformer architecture is equivalent to a numerical Ordinary Differential Equation (ODE) solver. Indeed, one can interpret deep learning as a parameter estimation problem of nonlinear dynamical systems [3] in which, the study of stability and convergence properties can be used to overcome the exploding and vanishing gradient phenomenon for example.

The goal of this project is divided into two parts. The theoretical approach, which consists in understanding the aforesaid transformer architecture and its translation into the ODE solver. After this initial step, derive and consider some improvements of the models used by considering other kind of ODE solvers or a more sophisticated interpretation in terms of the Koopman operator [4], for example. The second part will be the practical one, in which real simulations using the results obtained and the provided code will be carried out. More precisely, the main tasks to do are:

- Apply The ODE-Transformer model (https://github.com/libeineu/ODE-Transformer) to an OCR task and compare model performance (https://github.com/EauDeData/ODAOCR).
- Visualize the intermediate representations of the ViT architectures analyzed, both the previously trained models that are used as a reference, and the new models formulated in terms of EDOs.
- Apply dynamical system analysis techniques to analyse the behaviour of models formulated in terms of ODEs.
- Analyse whether the theoretical results obtained in the previous task are congruent with the results obtained when applied to the text recognition task.

This is a co-supervised project between the CRM and the CVC.

[1] Zhang, H., Wang, Y., Zhang, Z. (2021). *Continuous Self-Attention Models with Neural ODE Networks*. Proceedings of the AAAI Conference on Artificial Intelligence, 35(12), 10893–10901. https://doi.org/10.1609/aaai.v35i12.17445
[2] Li, B., et al. (2022). *ODE Transformer: An Ordinary Differential Equation-Inspired Model for Sequence Generation*. Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics, 8335–8351. https://doi.org/10.18653/v1/2022.acl-long.571
[3] Haber, E., Ruthotto, L. (2017). *Stable architectures for deep neural networks*. Inverse Problems, 34(1), 014003. https://doi.org/10.1088/1361-6420/aa9a90
[4] Geneva, N., Zabaras, N. (2022). *Transformers for modeling physical systems*. Neural Networks, 146, 272–289. https://doi.org/10.1016/j.neunet.2021.11.022