|  |  |
| --- | --- |
| **Format 3** : Thesis with research project within a research institute proposed by:  Research Group Digital Mathematicus (DIMA), Vrije Universiteit Brussel (VUB) | |
| Segmentation-Free Handwritten Text Recognition based on Joint Visual and Language Modeling | |
| GUIDANCE | |
| * For more information, please contact: | *Tan.Lu@vub.be* |
| * Academic promotor: (name) | Prof. Dr. Ann Dooms  [Ann.Dooms@vub.be](mailto:Ann.Dooms@vub.be) |
| * Supervisor and daily advisor(s) in the institute: (name) | Prof. Dr. Tan Lu  [Tan.Lu@vub.be](mailto:Tan.Lu@vub.be) |
| (name) | Prof. Dr. Tan Lu  [Tan.Lu@vub.be](mailto:Tan.Lu@vub.be) |
| * Number of students | 1 |
| CONTEXT | |
| Handwritten text recognition (HTR) has been primarily considered as a sequential data processing problem, where both classical pattern recognition models (e.g. hidden Markov model or HMM) and deep learning based techniques (e.g. recurrent neural networks or transformers) are being applied. However, to be able to apply these models, the input handwritten text images have to be pre-segmented to isolate individual text lines, which can then be processed by existing HTR models [1]. Unfortunately, the segmentation of handwritten text images is a trivial and error-prune process, which often renders HTR prohibitive in practical scenarios (e.g. thinking of the analysis of large amount of digitized manuscripts in the cultural heritage sector).  Recent advances in deep learning for both visual (e.g. using convolutional neural network or CNN) and language (e.g. using transformers) processing have promoted the investigation of possibly segmentation-free HTR models, where visual and language processing are combined such that HTR can be modeled as a 2-D computer vision problem [2]. While this immediately posts an interesting direction, many research questions are to be further investigated to develop and analyze such combined visual and language modeling. | |
| GOAL | |
| The goal of this master thesis is to investigate HTR modeling with an emphasis on:   1. design and implement segmentation-free HTR models by considering the combination of deep learning based visual and language processing; 2. investigate the performance of the implemented segmentation-free HTR models. | |
| METHODOLOGY | |
| The main methodology that will be applied in this thesis is deep learning, especially deep convolutional neural networks and transformers. | |
| PROFILE/REQUIRED SKILLS (e.g. rather theoretical / rather practical implementation, required knowledge (courses, methods, computer language(s), etc.) | |
| 1. good understanding in machine learning and neural networks 2. good skills in coding deep learning models using Python (e.g. using pytorch) | |
| COVID-19: In case additional restrictions are put into place by the government because of COVID-19, can the student continue the internship online? If adjustments to the project are needed in such a case, please specify these adjustments. | |
| The student can continue the thesis online if additional COVID-19 restrictions are implemented to prevent face-to-face discussions. | |
| REFERENCES | |
| 1. T. Plötz and G. Fink, Markov Models for Offline Handwriting Recognition: A Survey, International Journal on Document Analysis and Recognition (IJDAR), 2009 2. S. S. Singh and S. Karayev, Full Page Handwriting Recognition via Image to Sequence Extraction, arXiv, 2020. | |