Initial Research and Literature Review

Title	Methodology	Strengths	Limitations
TrOCR: Transformer-based Optical Character Recognition with Pre-trained Models (Minghao Li, et al. 2021)	•Transformer Architecture •Uses pre-trained CN and NLP models •Splits image into sequence of patches that are used as inputs	•SOTA Results •Uses pre-trained CN and NLP models, which take advantage of large-scale unlabeled data for image understanding and language modeling, with no need for an external language model •Does not require CNN for backbone, so image-specific biases are avoided	•Requires huge amount of data •Not suitable for low-resource languages (e.g Urdu), due to the nature of transformer architecture as it requires large amounts of data
LayoutLM: Pre-training of Text and Layout for Document Image Understanding (ACM, 2019)(Yiheng Xu, et al. 2019)	•Uses both text and document layouts for training •Joint training in textual and layout information •BERT is used as the backbone, and adds two new input embeddings: Positional and image embedding •Positional embeddings to capture relationship among tokens within a document	•SOTA Results •Takes into account both textual and layout information, which is beneficial for a great number of real-world document image understanding tasks such as information extraction from scanned documents	•Only works for English (will not work for Urdu, or multilingual use cases) •Needs a separate model for text extraction and localization, alongside the LayoutLM model itself, which is quite computationally heavy

Research Similarity

Title	Similarity
An attention-based method for offline handwritten Urdu text recognition (ICFHR, 2020)(Anjum and Khan 2020) [1]	• Urdu Language extraction, as well as we intend to localize language • This paper can act as our baseline for Urdu text extraction
A convolutional recursive deep architecture for unconstrained Urdu handwriting recognition (Neural Computing & application, 2021) [2]	• Urdu handwritten text extraction • FYP is about extracting bilingual text (Urdu and English) • Paper deals with the Urdu extraction part, which is a module in the Document text extraction.
A Computationally Efficient Pipeline Approach to Full Page Offline Handwritten Text Recognition (ICDARW, 2019) [3]	• We intend to extract text from a free form (full page) bilingual document (English and Urdu) • We also intend to work on a modular approach as proposed by this research paper
An online cursive handwritten medical words recognition system for busy doctors in developing countries for ensuring efficient healthcare service delivery (Scientific reports, 2022) [4]	• They work with handwritten, bilingual handwriting, which is precisely similar to our use case • Work in a similar demographic to ours, i-e, the authors use a developing country like Bangladesh for their case study
Multilingual handwritten numeral recognition using a robust deep network joint with transfer learning (Information Sciences (Elsevier),2021) (Fateh, et al. 2021) [5]	**• Very similar to what we intend to do, with both a multilingual approach, as well as considering Urdu in their problem as well, but limited to digits**
TrOCR: Transformer-based Optical Character Recognition with Pre-trained Models (Minghao Li, et al. 2021)	• This is essentially the application of TrOCR, but we intend to work on the handwritten text for both English and Urdu
LayoutLM: Pre-training of Text and Layout for Document Image Understanding (ACM, 2019)(Yiheng Xu, et al. 2019)	• Our research work also deals with entity extraction based on layout information, so there's a relation of this work in our research.