Project ID: CS24

Project Source: School of Computer Science

Project Title: Real-World VQA on PDF Documents

Project Description and Scope:

Document-based Visual Question Answering examines the document understanding of document images in conditions of natural language questions. The project will focus on making a document-based visual question-answering (VQA) system, PDF-VQA, to comprehensively examine the document understanding from various aspects, including document element recognition, document layout structural understanding as well as contextual understanding, and key information extraction.

Note that PDF-VQA would extend the current scale of document understanding that limits the single document page to the new scale that asks questions over the full document of multiple pages. The supervisory team will share the new graph-based VQA model that explicitly integrates the spatial and hierarchically structural relationships between different document elements to boost the document structural understanding.

Expected outcomes/deliverables:

The visual question answering system via PDF document understanding.

Specific required knowledge, skills, and/or technology:

Assumed/required knowledge and skills:

(Must have this knowledge before they select this

project) Python, Software Developers with the ability to design, develop, and code,

(Good to have but not essential) NLP, Machine Learning, Deep Learning

Fields that this project may involve:

NLP;Artificial Intelligence;

Resources provided by the client:

https://arxiv.org/abs/2304.06447