

Fine-tuning Vision Language Models to extract information from Bridge Technical Drawings

Task

This project aims to develop an automated system for extracting and interpreting information from BIM/bridge technical drawings using fine-tuned vision-language models (VLMs). Building on the work of Khan et al. (2023) [1], the key idea is to experiment with different open-source vision language models (VLMs) and fine-tune them to the specific domain of bridge engineering drawings and technical texts. The goal is to create a pipeline capable of accurately identifying and extracting information from 2D bridge technical drawings (such as structural elements, topological information dimensions, annotations etc.)

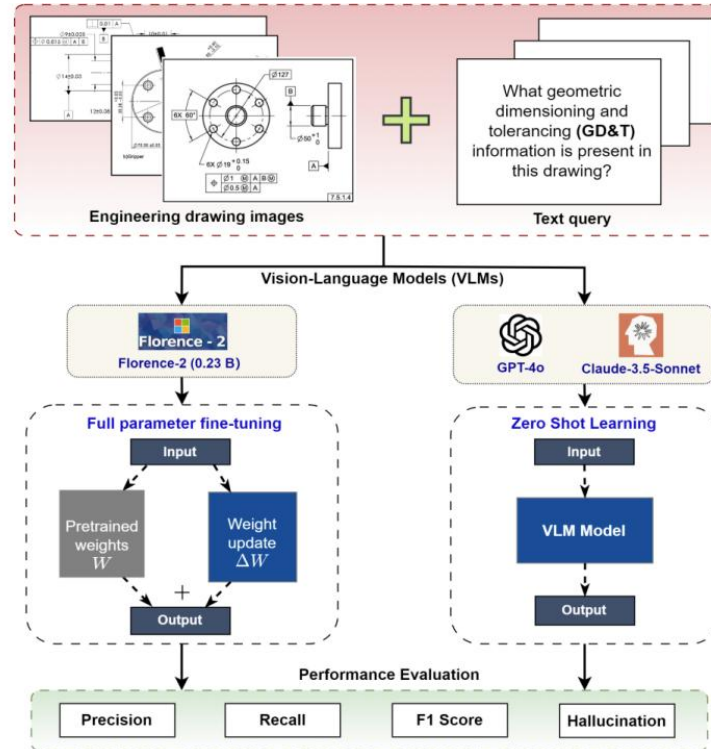


Figure 1: A general framework for information extractions from 2D engineering drawings. [1]

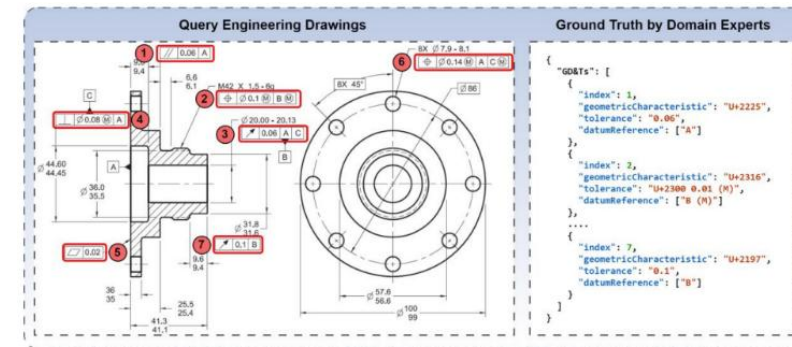
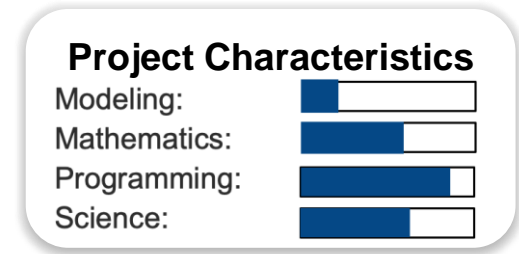


Figure 2: Example of dataset preparation (drawings with ground truth annotations in JSON format)

[1] Khan, Muhammad Tayyab, et al. "Fine-Tuning Vision-Language Model for Automated Engineering Drawing Information Extraction." arXiv preprint arXiv:2411.03707 (2024).