Sketch-Vision: Primitive Detection and Program Reconstruction

D1.2 Progress Report

Team JAXAXAX
Nikita Zagainov, Nikita Tsukanov, Said Kadirov, Dmitry Tetkin

October 7, 2025

Abstract

We extend raster-to-program modeling to hand-drawn sketches with engineering annotations. This interim deliverable documents (i) a lightweight synthetic dataset generator with JSON annotations, (ii) visualization and evaluation utilities, and (iii) updated repository documentation. We outline alignment with CAD2Program-style VLMs and next steps.

1 Overview

Our goal is to detect digits, arrows, dimensions, radii, and geometric primitives in noisy sketches, producing a structured representation suitable for downstream CAD. We follow the encoder–decoder paradigm (ViT encoder + LM decoder) summarized in our prior notes (see ‘docs/sketch-vision-eng.pdf‘).

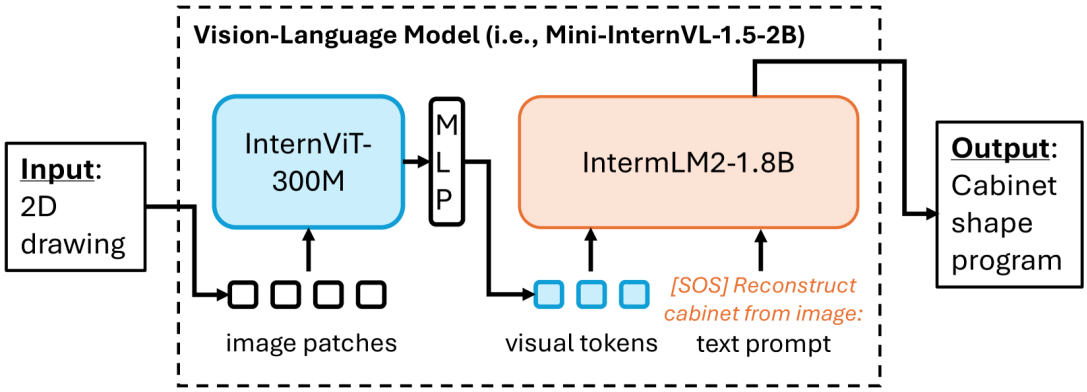

Figure 1: High-level multimodal architecture (reference figure in repo).

2 Repository Updates in D1.2

2.1 Synthetic Dataset Generator

We added ‘preprocessing/generate$_s$*ynthetic.py*‘*thatrendersrectangles*, *circles*, *andlinesegmentswithbasicdime*

2.2 Visualization and Evaluation

‘preprocessing/visualize$_a$*nnotations.py*‘*overlaysboundingboxesandlabelsontopofimages*.‘*evaluation/metrics.p*

2.3 Documentation

‘README.md‘ now includes a Quickstart showing how to generate a small dataset, visualize a sample, and get stats.

3 Quickstart (Reproducibility)

```
python preprocessing/generate_synthetic.py --output-dir dataset/synthetic --num-samples 100
python preprocessing/visualize_annotations.py \
  --images-dir dataset/synthetic/images \
  --annotations-dir dataset/synthetic/annotations \
  --name sketch_00010 --output dataset/synthetic/vis
python evaluation/evaluate_synthetic.py \
  --annotations-dir dataset/synthetic/annotations \
  --splits dataset/synthetic/splits/train.txt
```

Generated visualizations can be included in the appendix; see ‘docs/example.png‘ for style reference.

4 Planned Work

- Expand primitive set (arrows, dimensions-as-text tokens) and OCR integration.
- Train a detector baseline (ViT/DETR) on synthetic data; add metrics (precision/recall/F1).
- Integrate encoder–decoder path for program reconstruction; evaluate sequence accuracy.

Acknowledgments

We thank Innopolis University for support and compute.