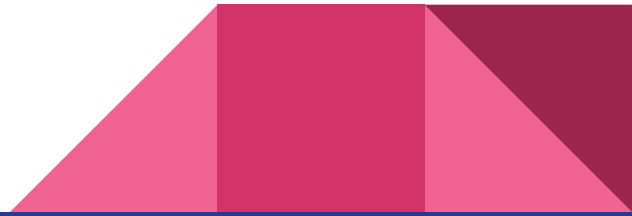
Multi-modal Deep Learning for Automated Schematic Analysis

Vojtěch Sýkora

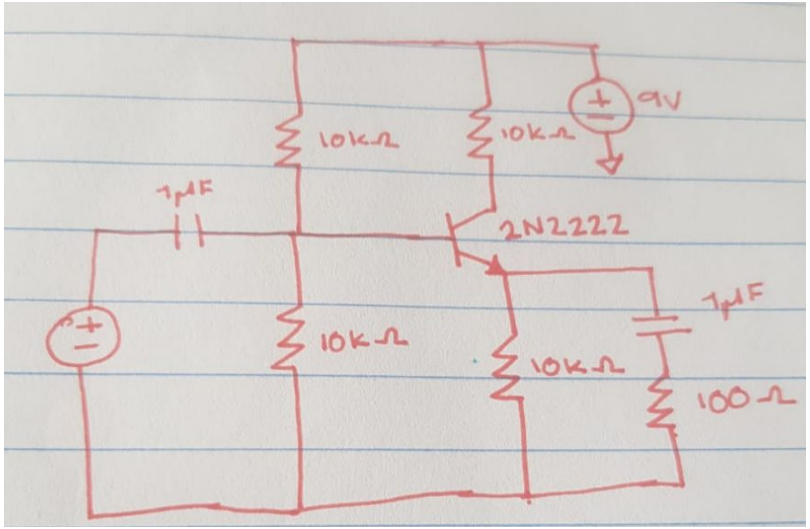
Motivation

- Missing good digitization tools for electric circuit diagrams
- Accurate and automated parsing of images of diagrams could save a lot of time and money.
- Focus: Raspberry Pi devices are used everywhere
- Goal: increase performance of an existing extraction software and bridge gap between hand-drawn (CGHD) and computer-generated (RPi) schematics with a robust pipeline.

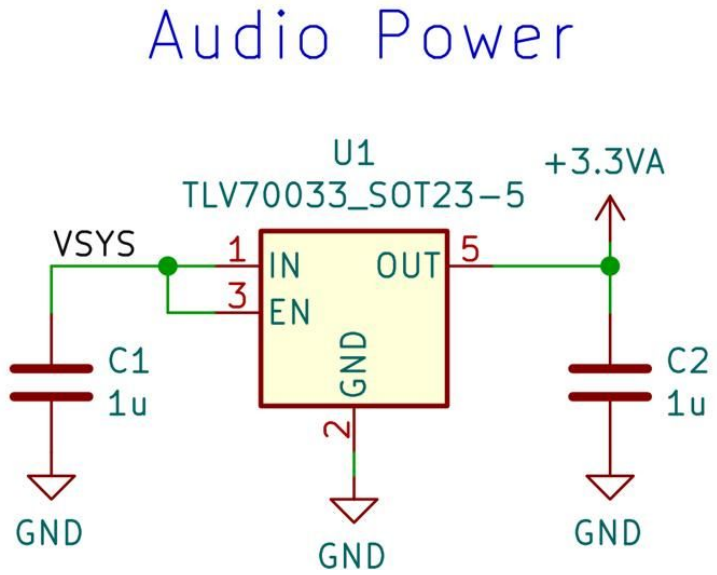


Datasets & Domain Gap

Train, Val → Domain Transfer → Test



CGHD Dataset



Raspberry Pi - RPi Dataset

Baseline Modular Pipeline

- Faster R-CNN for object detection - 59 classes
- Rotation, segmentation, OCR -> assemble graph of the diagram

Table 2.2.: CGHD and RPi Datasets

Category	CGHD	RPi
Annotated Raw Images / Files	3,173	22
Bounding Box Annotations	245,962	1,675

Faster R-CNN Enhancements (Object Detection)

- Focal Loss → tackles severe class imbalance
- GIoU Loss → tighter bounding-box regression
- Efficient Channel Attention (ECA)→ distinguish similar symbols
- Dilated Convolutions → enlarge receptive field for larger symbols
- Training hyper-params:
  - Max predicted boxes
  - Classics - lr, batch size …
  - Image transformations - RandomErasing …

Preliminary Positive Results (+10% mAP)

Table 2.2.: Preliminary Results on CGHD Validation Set (mAP)

Configuration	**mAP (%)**
Baseline	44.3
Focal Loss	54.2
GIoU Loss	54.2
Efficient Channel Attention (ECA)	53.3
Dilated Convolutions	52.2
Focal + GIoU	**54.3**
Focal + GIoU + Noise	53.6
Focal + GIoU + Erasing	53.8

Vision-Language Models (Molmo-7B-D)

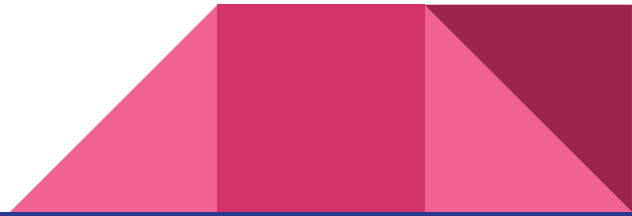
- Metric 1: symbols found in image
- Metric 2: bbox locations of found symbols

Findings so far:

- Smaller schematics - some understanding of locations and symbols
- Bigger schematics - no understanding of locations, some understanding of symbols

Next Steps

- Faster R-CNN: Test image transformations on best model so far (50% augmented, 50% original)
  - Evaluate best model on RPi dataset against baseline
- Explore more VLM prompt engineering
  - Few-shot prompting
  - Chain-of-thought



Sources

- F. Thoma, J. Bayer, and Y. Li, **CGHD**: A public ground-truth dataset for handwritten circuit diagram images, 2021. arXiv: 2107.10373 [cs.CV].
  - Image from slide 3 from “drafter 1/C2_D2_P2.jpg”
- J. Bayer, L. van Waveren, and A. Dengel, **Modular graph extraction for hand-written circuit diagram images**, 2024. arXiv: 2402.11093 [cs.CV]
- Raspberry Pi Foundation, **Raspberry pi datasheets**, https://datasheets.raspberrypi.com/, Accessed: 2025-06-01, 2024.
  - Image from slide 3 from rp2040/vga-for-pico-w-schematics.pdf
- S. Ren, K. He, R. Girshick, and J. Sun, “**Faster r-cnn**: Towards real-time object detection with region proposal networks”, in Advances in Neural Information Processing Systems, 2015, pp. 91–99.