Wentao Yao

📞 618-739-8246 | ✉ wentao.h.yao@gmail.com | 🔗 Wentao Yao | ⌂ https://harrygyao.github.io

EDUCATION

Northwestern University   Evanston, IL
Master of Science in Artificial Intelligence; GPA: 4.0/4.0   *Sept. 2021 – Dec. 2022*
The Ohio State University   Columbus, OH
Bachelor of Science in Computer Information Science; GPA: 3.5/4.0   *Aug. 2017 – May 2021*
Minor in Mathematics

TECHNICAL SKILLS

Languages: C, C++, Java, Python, SQL, MATLAB, and Assembly,
Machine Learning: CNN, R-CNN, YOLO, GAN, RNN, LSTM, Transformer, and DQN
Frameworks & **Libraries**: PyTorch, TensorFlow, MySQL, Postgres, and SpringBoot
Tools & **Other:** Git, Jira, Docker, AWS, PCIe and NVMe

EXPERIENCE

Samsung Semiconductor Inc.   Jan. 2023 – Present
Firmware Engineer   *San Jose, CA*

- **GEN-5 SSD:** Developed and maintained firmware for Samsung Semiconductor's Gen-5 SSD products based on the **NVMe** specification.
- Optimized firmware performance and reliability through extensive testing, validation, and performance analysis.
- Contributed to the development of new features and enhancements for Samsung's SSD products, including firmware optimization and innovation Worked with the quality assurance team to ensure high-quality firmware and timely product releases
- Collaborated with cross-functional teams, including hardware engineers, system architects, and software developers, to deliver high-quality firmware solutions.

Precise Software Solutions Inc.   Jun. 2022 – Aug. 2022
Data Scientist Intern   *Rockville, MD*

- **Document A.I.:** Detected and accurately annotated figures from images of page-based documents. Build prediction model based on the **Faster R-CNN** model pre-trained on the **PubLayNet**. Participated in the design of a post-processing pipeline for improving performance.
- Designed a new evaluation metric that replaced the **IOU(intersection over union)** to measure the integrity of the figure. The post-processing increased the accuracy of annotation for the detected figures **from 0.9623 to 0.9926** on the **DocBank** dataset.
- Designed and implemented an object-orientation **end-to-end machine learning pipeline** to control the workflow and handle cloud service sequences, including **AWS Textract, Google Vision, and Adobe PDF AI**.
- Achieved reconstruction of PDF file from the **scanned image** into a **searchable document**.

The Home Depot   Mar. 2022 – Jun. 2022
Machine Learning Practicum Student   *Evanston, IL*

- **Point-of-sale (POS) system:** Created a **deep learning pipeline** to recognize and locate products when customers scan at the self-checkout based on **YOLO-V4 model**.
- Used pre-trained **Mask RCNN** model and **GrabCut** algorithm to get segmentation mask and remove occlusion on the target product.

PROJECTS

- **Covid-19 Twitter Analysis:** Used sentiment analysis on the topic of the Covid-19 vaccine to find the trend of public opinion on Covid-19 over time. Tech: **pandas, NLP, BERT, Graphic Analysis** (Mar. 2022 – Jun. 2022)
- **Image Style Transfer:** Achieved the transfer of Monet's and Van Gogh's painting styles to each other by training a cutting-edge CycleGan model. Designed a new pipeline to solve the underfitting problem caused by the small dataset. Tech: **Python, GAN, U-net, TPU, TensorFlow** (Jan. 2022 – Mar. 2022)
- **Citizen's Police Data Project:**Analyzed Police Database with Postgres. Generated visualized complaint reports against the police on Tableau after filtering out invalid data via data cleaning. Tech: **Postgres, Tableau, Data Cleaning, Graphic Analysis, NLP**(Sept. 2021 – Dec. 2021)