Shen-En Chen (Andrew Chen)

Atlanta, GA | +1 (404) 376 0019 | Hsinchu, Taiwan | +886 975596112
achen353@gatech.edu | www.linkedin.com/in/achen353 | achen353.github.io

EDUCATION

Georgia Institute of Technology — Atlanta, GA

- M.S. in Computer Science (Specialization: Machine Learning) — Aug. 2021 – Dec. 2022
- B.S. in Computer Science (Threads: Intelligence & Info Internetworks), GPA 3.96/4.0 — Aug. 2018 – May 2021

SKILLS

- Programming/Markup Languages: Python, Java, C++, C, HTML, CSS, JavaScript, LaTeX, SQL
- Libraries/Frameworks/Platforms: Numpy, Scipy, Pandas, scikit-learn, PyTorch, Tensorflow/Keras, OpenCV, NLTK, Spacy, genism, gRPC, Kubernetes, Docker, Google Cloud Platform (GCP), Azure Cognitive Services
- Relevant Coursework: Natural Language Processing, Machine Learning, Computer Vision, Intro to Info Security

EXPERIENCE

GliaCloud Co., Ltd. — Taipei, Taiwan

AI Intern — Apr. 2021 – Present

- Fine-tuned and deployed to **Google Kubernetes Engine (GKE)** a custom multilingual **BERT-Viterbi keyword model** to improve performances on French and Indonesian by 10% and 13.6%, respectively, while retaining 97% of the performances on already supported English, Mandarin Chinese, Japanese, Korean, and Vietnamese.
- Leveraged **knowledge distillation** and **input embedding reduction** to perform **model compression** on the aforementioned **transformer model**, reducing the model size by 48% with 97% of performances preserved.
- Evaluated Google Cloud and Azure **Automatic Speech Recognition (ASR)** using on **word error rate (WER)**, **match error rate (MER)**, **word information preserved (WIP)**, and custom-defined interval-based **precision and recall**.
- Researched and tested on different **voice activity detection (VAD)** algorithms including energy-based **frequency domain** filtering and support vector machine (SVM) and convolutional neural networks (CNNs) on various **acoustic features**.

Medical Informatic Research and Genetic Elucidation Lab, National Taiwan University — Taipei, Taiwan

Summer Research Intern — May 2019 – Aug. 2019

- Built a **facial recognition** program using **OpenCV** and convolutional neural networks.
- Designed a **machine learning** classification model for 5 common lung tumor types using ensemble one-vs-one **support vector machine (SVM)** classifier.
- Applied **3D** residual **convolutional neural networks**, using **Keras** and **scikit-learn**, on augmented Lung Image Database Consortium image collection (LIDC-IDRI) to classify benign and malignant lung tumors and achieved an accuracy, sensitivity, and specificity of **97.23%**, **95.54%**, and **98.12%**, respectively.

PROJECTS

Sentiment Analysis and Topic Modeling on COVID-19 Vaccine Tweets — Atlanta, GA

Team Lead — Jan. 2021 – May 2021

- Performed **exploratory data analysis** and built a **data preprocessing** script with **NLTK** and **spacy** libraries.
- Applied **Latent Dirichlet Allocation (LDA)** and **Hierarchical Dirichlet Process (HDP)** to uncover inherent tweet topics.
- **Led** a team of 5 to perform **sentiment analysis** with sentiment dictionaries and **bi-directional LSTMs**.

Taiwanese Traffic Object Detection — Hsinchu, Taiwan

Personal Project — Dec. 2020 – Jan. 2021

- Trained and fine-tuned **Darknet YOLOv4 Tiny** at different resolutions, learning rates, and momentum.
- Achieved an **87.5% mAP@0.5** at about **18 to 23 average FPS** with Nvidia Tesla P100 GPU.

Proper Mask Wearing Detection and Alarm System — Hsinchu, Taiwan

Personal Project — Dec. 2020 – Jan. 2021

- Performed transfer learning on **MobileNet V2** using Keras, OpenCV, and **Google Compute Engine (GCE)**.
- Designed and deployed a **real-time detection web app** on **Google App Engine** using **Dash** framework.

ITS-Chatbot — Atlanta, GA

Team Lead & Member — Jan. 2020 – May 2021

- Built a **retrieval-based chatbot** model using **cosine similarity** on **fastText** word embeddings with **TF-IDF**.
- Incorporated and tested a **BERT** and an **ELECTRA question-answering** model using the **Hugging Face** library.
- **Led** a team of 4 to train and compare **genism LDA**, **Mallet LDA**, **tomotopy LDA**, and **CorEx** topic modeling algorithms on DSP First textbook data and improve them with key term mappings.