

Yuan (Jessica) Huang

Email: yuanhuan@usc.edu Tel: 213-519-2954

Education Background

- **University of Southern California (MS in Electrical Computer Engineering – machine learning)** GPA:3.5/4.0
- **East China University of Science and Technology (BS in Information Engineering)** GPA:3.5/4.0
Second Scholarship for Academic Excellence at each year (2016 – 2020)

Work Experience

- Allegro Lab at USC** **Research Assistant (volunteer)** **Apr.2022-now**
- Contributed to several projects in the fields of VLN (Vision Language Navigation) and NLP (Natural Language Processing) utilizing different deep learning models in Python, experienced in language data processing and analysis.
 - Monthly group meeting presentations, weekly project meetings with professor. Second author of a paper submitted to EMNLP 2023.
- Beijing Sunwayworld Technology Corporation, Ltd** **Software Engineer** **Apr.2020-Nov.2020**
- In charge of the daily maintenance of Oracle database, including backup and recovery, log analysis, data migration, etc.
- Shanghai Advanced Avionics Corporation, Ltd** **Hardware Engineer (intern)** **Jul.2017-Aug.2017**
- Completed the aviation test, equipment cabinet project and the cabinet-related inspection work.
 - Drew PCB board by Protel, drew circuit board module with AutoCAD, tested the board functions.

Projects Experience

- Scale laws analysis on synthetic data augmentation** **May.2023-Aug.2023**
- Analyze the scale laws of synthetic data generation using models with different size of parameters. Generate data with different models and compare the perplexity of the outputs.
 - Using open source LLMs (such as Pythia) with in context learning and finetuning. (Full/ Adapter)
 - Comparing in-context learning, few-shot finetuning and synthetic data finetuning.
 - Steer project direction and independently handle coding task.
- Synthetic data augmentation on different models** **Oct.2022-May.2023**
- Worked in natural language processing (NLP) using Python.
 - Collect raw data and do data cleaning then establish the dataset. 1)Use different model to learn from the target data. (1.1) learn directly (1.2) construct question answer and false answer separately for QA 3) GPT-based generation (prompting).
 - Achieved 10% accuracy increasement from the baseline. Although unable to surpass the GPT-4's performance, leading to a strategic shift in project direction.
- Vision-language navigation based on noisy data augmentation** **May.2022-Feb.2022**
- Evaluated the robustness of wide-used multi-modal reasoning systems under both multi-image and single-image setups.
 - Analyzed VLN agents by adding noise to the pretraining data. Demonstrated word ordering and trajectory matching does not matter much in VLN pretraining. Demonstrated the data quantity is more important than data quality in VLN pretraining. Proposed a unigram + landmark method to lower the cost of data creation and augmentation. Designed a nonsense-data augmentation method for effective VLN pretraining.
 - My co-2nd-auther work has been sent to the EMNLP 2023 conference.
- Audio recognition project based on sound dynamics signals (As team leader)** **Sep.2018-Oct.2019**
- Participated in the College Students' Innovation and Entrepreneurship Competition and got school awards
 - Used GTZAN dataset and Gammatone filter for feature extraction, used RMDL model. (acc 89%)
- Piano audio recognition based on MATLAB** **Sep.2018-Dec.2018**
- Modeled on frequency characteristics in piano. Divide the entire piano audio into frames by time domain analysis and endpoint detection. Recognized the notes or chords in the frequency domain.
- Communication system simulation based on MATLAB platform** **Sep.2017-Dec.2017**
- Learned to design communication system and completed the simulation of the spread spectrum system.

Publications

- Tian Yuhao, Huang Yuan. Research on Image Generation based on Adversarial network. [J]. *Electronic Technology & Software Engineering*, 2019(22):68-69[2019-12-06]

Skills

Computer skills:

1. Programming language: Python, JAVA, C/C++, MATLAB.
2. Proficient Python packages: TensorFlow, PyTorch, Keras, Pandas, Matplotlib and more.
3. LLMs: I have worked with large language models such as BERT, RoBERTa, T5, OpenAI GPT models, etc.
4. Other: Linux and CUDA

My GitHub link is [here](#). My LinkedIn link is [here](#). I am familiar with AWS (EC2,etc) and google cloud platform (GCP).

About me

- Highly motivated, self-starting and passionate person
- Keep updated with the newest tech.
- Easygoing and good at team working.

Personal Interests:

- Singing, hiking, and reading.