

# CHEN-KANG LEE

☎ +1 (858)-500-2105     [Chen-Kang Lee](#)     [ChenKangLee](#)    ✉ [chl126@ucsd.edu](mailto:chl126@ucsd.edu)

## EDUCATION

- University of California San Diego** Sep 2021 - Dec 2022 (Expected)  
*M.Sc. in Computer Science and Engineering (GPA: 3.84/4.0)* La Jolla, CA, USA
- Highlighted Coursework: Introduction to Robotics, Algorithm Design and Analysis, Advanced Compiler Design
- National Tsing Hua University** Sep 2015 - Jun 2019  
*B.S. in Computer Science (GPA: 3.82/4.0, Academic Achievement Award)* Hsinchu, Taiwan
- Highlighted Coursework: Operating Systems, Data Structures, Embedded Systems, Parallel Programming

## SKILLS

**Programming Languages** C/C++, Python, Javascript/Typescript, C#, SQL  
**Tools & Framework** Pytorch, Scikit-learn, NumPy, Pandas, Tensorflow, OpenCV, Docker, Git, React, ROS  
Unity3D, Apache Beam, Apache Airflow, MapReduce, MPI, Google Cloud Platform

## PROFESSIONAL EXPERIENCE

- TechOps Intern** Jun 2022 - Present  
*Pyramid Systems Inc.* Fairfax, VA, USA
- Developing an automatic machine learning pipeline on Google Cloud Platform to identify natural disaster hot zones from FEMA disaster data and classify banks by their risks.
  - Implementing an automatically scalable distributed data preprocessing pipeline using **Apache Beam**.
  - Implementing and training machine learning model on **Vertex AI** and automation using **Apache Airflow**.
  - Creating a full stack web application using **React** front end to present the result of the pipeline, and deployed the application server on **Google App Engine**.
- Research Assistant** Jun 2020 - May 2021  
*Institute of Information Science, Academia Sinica* Taipei, Taiwan
- Developed H-FND, an iterative **reinforcement learning** framework that achieved state-of-the-art performance on relation extraction on the SemEval distantly supervised dataset with false negative noise.
  - Utilized the proposed framework to automatically extract new knowledge from un-annotated news articles. [[Demo](#)]
  - Developed and maintained the relation extraction functionality of the lab's CoreNLP pipeline. [[Demo](#)]
- Data Scientist Intern** Jan 2019 - Mar 2019  
*SkyEyes GPS Technology Co., Ltd.* Taichung, Taiwan
- Implemented MobileNet in **Pytorch** to detect mis-classification in the driver assessment software with an 88% accuracy.
  - Preprocessed images taken by dash cams of the driver assessment software to generate training data for the deep learning classification model.

## PROJECTS

- Rain-induced Landslide Prediction** Jan 2020 - Apr 2020  
*Project as Research Assistant at FCU AI Research Center*
- Preprocessed historical satellite landslide data, geographic data and rainfall data to predict rain-induced landslides in the Chenyulan river watershed.
  - Co-developed a **graph-based sequential model** that utilizes **generative adversarial networks (GANs)** to combat the lack of training data. The model achieved 85% accuracy (0.72 F1-score) predicting landslides in test.
- Rental Bike Flow Prediction** Oct 2019 - Jan 2020  
*Project as Research Assistant at FCU AI Research Center*
- Implemented a **multi-graph GCN model** on NYC Citibike rental flow prediction. The model was able to utilize the spatial relation of the rental stops and outperform the baseline LSTM models by **25%** (reduction in RMSE). [[Github](#)]
  - Adapted the graph-based model to the Taiwanese bike rental service, where we incorporated the **Transformer** architecture to better capture the more intense fluctuation of bike flows.

## PUBLICATIONS

Jhi-Wei Chen, Tsu-Jui Fu, **Chen-Kang Lee**, Wei-Yun Ma  
"H-FND: Hierarchical False-Negative Denoising For Distant Supervision Relation Extraction" [[Link](#)]  
*59th Annual Meeting of the Association for Computational Linguistics (ACL), 2021 (Findings)*