Pinheng Chen

New Graduate of IT in Software Development and Data Science

pionchen.stayhunsgry@gmail.com 0493416188 Sydney, Australia Temporary Graduate Visa (Subclass 485)
https://www.linkedin.com/in/pionstayhungry https://www.github.com/PionC

EDUCATION AND WORKING EXPERIENCE

Master of Information Technology (Major in Artificial Intelligence, Data Science and Engineering)

University of New South Wales **Sydney, Australia** **Feb 2022 - Feb 2024**

Software Developer - Intern

South Digital Technology Co., Ltd. **Guangzhou, China** **Jul 2020 - Dec 2021**

Bachelor of Geomatics

Shandong University of Science and Technology **Qingdao, China** **Sep 2016 - Jun 2020**

TECHNICAL STRENGTH

Front-end: JavaScript, TypeScript, React.js, Next.js, Angular, D3.js
Back-end: C/C++, Python, Node.js, Clingo
Deep Learning: Pytorch, Tensorflow, Scikit-Learn, Keras
Data Analysis: Numpy, Pandas, Matplotlib
Big Data: Hadoop, Spark, Google Dataproc
Database: PostgreSQL, Oracle, MongoDB
Game Development: Ureal Engine, Blueprint, Blender
Image Processing: OpenCV, Blender
Cloud Services: Google Cloud, Amazon Web Services, Mircosoft Azure
GIS/Geomatics: ArcGIS, AutoCAD, Surveying, GNSS
Agile Project Management: Jira, Kanban

HIGHLIGHT PROJECTS

Software/Web Development

- **Site Selection and Routing of Offshore Wind Farm**
  Project Overview: The project aimed to help the local government select the optimum site location and route of offshore wind farms so that the government can minimise the cost, maximise the electricity generation, decrease the risk of being destroyed by disasters and protect the marine ecological environment.
  Main Job:
  - Virtual machine set up (VMware Workstation)
  - Design and create interfaces (Node.js)
  - Database installation, set up and maintenance (Oracle)
  - Data fusion, integrating land and ocean geography data, land value, disaster, biodiversity data and wind resource data
- **Marine Disaster Monitoring Platform**
  Project Overview: The project aimed to help the local government monitor marine disasters, it integrates the data including seismic wave data, sea level data, climate data, etc. Once the marine disaster occurs, the platform can trigger the warning and give the corresponding solution which has been prepared in advance.
  Main Job:
  - Create a sample for software copyright application and help the company get the copyright of software in China
  - Design and create interfaces (C++, QT Creator)
  - Database installation, set up and maintenance (PostgreSQL)
- **3D Submarine Terrain Generation System**
  Project Overview: The project aimed to help the local government build up a system that can generate 3D submarine terrain with sonar data and multibeam echosounder data. The 3D submarine terrain can be zoomed in/out and change the perspective position. The location of a specific point and the distance between two points can be measured by the provided tools.
  Main Job:
  - Data preprocessing and filtering (Python)
  - Measurement adjustment (Python)
  - Build up interfaces (Node.js)
  - Database set up and maintenance (Oracle)
- **Smart PDF Reader**
  Project Overview: The project aimed to build a website that is able to provide users with a smart PDF reader. The registered users can upload their PDF files to the cloud storage space and manage their own files. Also, they can use the cross-reference function when they are reading a PDF file. Moreover, almost all of the citations and references of a paper can be shown visually on a page.

Main Job:

- Front-end Part (React.js, Tailwind CSS)
  - Customising PDF.js so that some functions can be changed (PDF.js)
  - Cross-reference (PDF.js)
  - Visualize the citations and references of the corresponding paper as nodes in a pseudo-random distribution way (D3.js)
- Back-end Part (Node.js)
  - Design and create interfaces
  - User's identity authentication and private information management and encryption
  - MongoDB and Google Cloud Buckets setup and management (MongoDB, Google Cloud Buckets)
  - PDF files uploading, deleting, renaming, altering, etc.
  - Extract the citations and references of the corresponding paper (Semantic Scholar)
  - Randomize the position of citations and references nodes with divide and conquer

Deep Learning

- **Pedestrians Detection on Self-automated Cars**
  Project Overview: The project aimed to apply a model based on YOLOv5 to detect pedestrians from the perspective of self-automated cars, optimizations are implemented to increase the accuracy or the efficiency of pedestrian detection
  Main Job:
  - Cloud server setup and configuration
  - Data set preparation and filtering
  - Data preprocessing, such as histogram equalization, white balance, geometric transformations, mosaic, etc. (Numpy, Matplotlib, Pandas, Pytorch)
  - Combine the basic YOLOv5 network structure with MobileNet and SE/ECA/SPD attention mechanicals (YOLOv5)
- **The CSIRO Crown-of-Thorn Starfish Detection**
  Project Overview: The project is actually a Kaggle competition that challenges the International Machine Learning community with the task of COTS detection from these underwater images. (More details can be available from The CSIRO Crown-of-Thorn Starfish Detection Dataset
  Main Job:
  - Cloud server setup and configuration
  - Data preprocessing, such as dark channel prior(a haze removal method), histogram equalization, white balance, geometric transformations, mosaic, etc. (Numpy, Matplotlib, Pandas)
  - Combine the basic YOLOv5 network structure with MobileNet and SE/ECA/SPD attention mechanicals (YOLOv5, Tensorflow)
  - Build up Faster-RCNN structure
- **Generating gender-ambiguous voices for privacy-preserving speech recognition**
  Project Overview: The project aimed to generate gender-ambiguous voices in order to protect users' private information through the model obtained by GenGAN
  Main Job:
  - Dataset preparation, data preprocessing and filtering
  - Build up GenGAN structure (Pytorch, Numpy, Matplotlib)
  - Train the model
  - Compare the result with the paper

Big Data

- **Detection of Popular and Trending Topics From News and Articles**
  Project Overview: The project aimed to detect popular and trending topics from news articles by performing text data analysis over a dataset of Australian news from ABC using MRJob
  Main Job:
  - Compute the weights of each term regarding each year in the news articles dataset (Python, MapReduce, MRJob)
  - Find out the most important terms in each year whose weights are larger than a specific given threshold
- **Detection of Top-k Most Frequent Co-occurring Term Pairs From News**
  Project Overview: The project aimed to detect popular and trending topics from news articles by performing text data analysis over a dataset of Australian news from ABC using both RDD and DataFrame APIs of Spark with Python
  Main Job:
  - Compute the weights of each term regarding each year in the news articles dataset (Python, MapReduce, PySpark)
  - Find out the top-k most important terms in each year.
- **Finding Similar News Article Headlines**
  Project Overview: The project aimed to find all similar news article headline pairs from news articles by a more efficient method of calculating the Jaccard Similarity Function over a dataset of Australian news from ABC using PySpark
  Main Job:
  - Find all similar news article headline pairs across different years (Python, MapReduce, PySpark)
  - Use the Jaccard Similarity Function to compute the similarity
  - Run the code in Google Dataproc
  - Apply PPJoin to reduce the time complexity of generating pairs