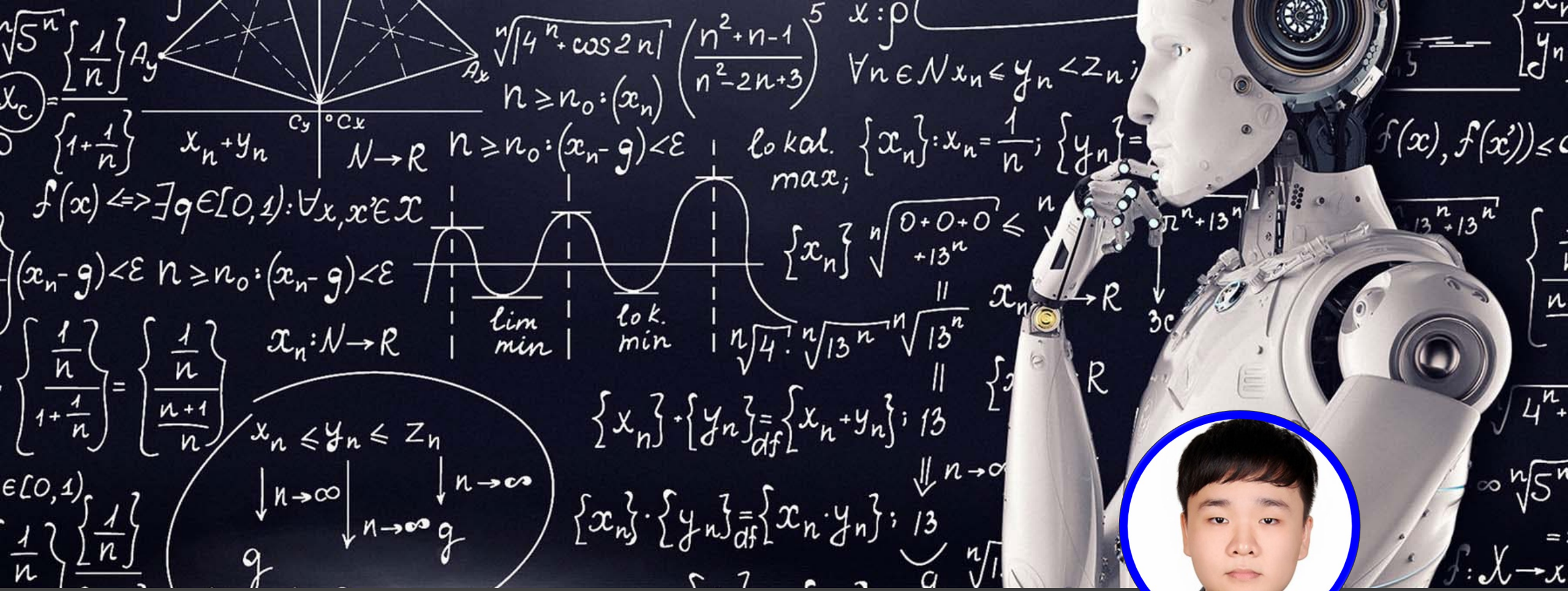


Deep learning Based Image Recognition
Course Logistics

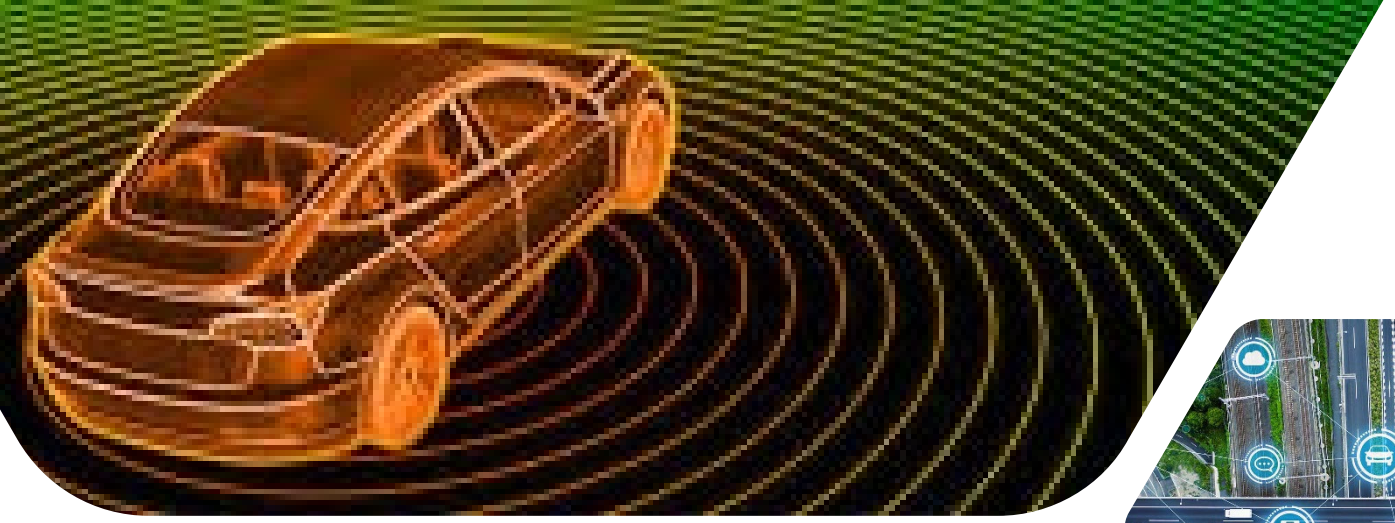
Instructor: 鍾昕燁

About Me

鍾昕燁 Sin-Ye Jhong

- **Current Positions**:
  - **Assistant Professor**, Dept. GIMT, NTUST
- **Work Experiences**:
  - **Assistant Professor**, Dept. AI, Tamkang University (TKU).
  - **Research Fellow**, Smart Electric Vehicle (SEV) Research Center, NTUST. (Recognized for **Outstanding R&D Substitute Service**)
  - **Visiting Scholar**, Dept. EE, University of South Florida (USF), Florida, USA (Supported by **NSTC Postdoctoral Research Abroad Program**)
  - **Full-time Computer Vision Intern**, Real Touch Corp. Ltd.

- **Research Areas**:
  - Computer Vision and Image Processing
  - Machine/Deep Learning
  - Reinforcement Learning
  - Multimodal Fusion
  - Generative AI
  - Agentic AI
  - Smart IoT
  - Intelligent Unmanned Vehicles
  - Smart Manufacturing
  - Intelligent Transportation



Research Achievements

- **Journal Publications**: 20
- **Conference Papers**: 21
- **Patents**: 10
- **Competition Awards**: 47

Course Logistics

- Instructor: 鍾昕燁 Sin-Ye Jhong
- Time & Location:
  - Tuesday, 13:20-16:10, TR-615
- Email: sinyejhong@mail.ntust.edu.tw
  - To ensure timely response, please include **"[DLIR2025]"** in the email subject
  - Leaving a message on Moodle is also an option, but responses may not be immediate

Always check the up-to-date information from Moodle

Course Goal

- **Problem-Based Learning (PBL):**
  - This is an advanced, hands-on course focused on independent research and implementation.
  - **Goal:** To research, implement, and analyze cutting-edge papers in image recognition.
- **Core Task:**
  - **Research & Implementation:** You are required to independently research and implement **four academic papers** (**2** top-tier conference, **2** Q1 journal)
  - Presentations: You will deliver **four** presentations in total:
    - ✓ **Two 10-minute presentations** on your paper research (during the semester).
    - ✓ **One 20-minute** midterm project presentation.
    - ✓ **One 20-minute** final project presentation.

Key Requirements for Success

- **Skills:**
  - Proficiency in Python, PyTorch, Git, Docker, ROS, deep learning and image processing fundamentals is required
- **Hardware (Mandatory):**
  - A high-performance PC is essential. A GPU with **at least 16GB of VRAM** (e.g., RTX 3090, 4070Ti Super,4080, 4090, 5070Ti, 5080, 5090) is required for model training
- **Time Commitment:**
  - This course is demanding. Expect to invest **at least 20 hours per week**

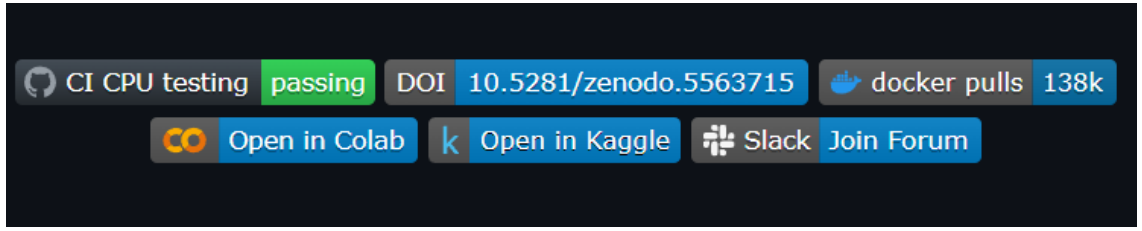
Research Paper Requirements & Deadlines

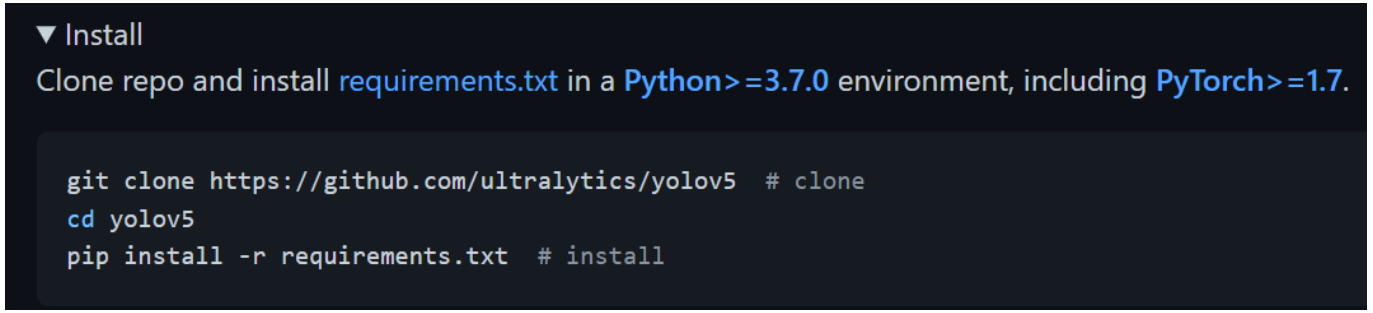
- **Core Task:** Read and implement 2 top-tier conference papers & 2 Q1 journal papers (from 2022-2025)
- **Deliverables (For Each Paper):**
  - **Written Report:** A detailed analysis of the paper's contributions, methods, and experiments, with clear explanations of your code
  - **Implementation Code (with Docker)**: Your code must successfully reproduce the key results described in the paper
- **Schedule & Deadlines:**
  - Paper Topic Confirmation: Paper 1 (9/9), Paper 2 (9/30), Paper3 (10/28), Paper 4 (11/18)
  - Final Submission: Paper 1 (9/23), Paper 2 (10/14), Paper3 (11/11), Paper 4 (12/2)

Grading: Your performance on these tasks will directly impact your participation, midterm, and final grades.
Time Commitment: Be prepared to invest at **least 20 hours per week** on this work. **Please evaluate your schedule and capabilities carefully**.

Example: YOLOv5 as a Standard for Good Projects

- Yolov5(ultralytics): https://github.com/ultralytics/yolov5:





- **Your Project Requirements:**
  - GitHub: All of your project code must be hosted on GitHub
  - README.md: You must write a detailed README.md file
  - Docker: A Dockerfile is mandatory to ensure your project is reproducible and can be easily tested by the TA.

Grading Policy

- **Participation & Paper Reports: 35%**
  - Includes class engagement and **four** paper implementation reports.
  - HW0 is 15% of this grade
- **Midterm Project: 30%**
  - Evaluated on Live Demo (20%), Performance Metrics (20%), your 20-min Oral Presentation (30%), and a Written Report (30%)
- **Final Project Project: 30%**
  - Evaluated on Live Demo (20%), Performance Metrics (20%), your 20-min Oral Presentation (30%), and a Written Report (30%)

All project code must be successfully demonstrated on the TA's computer.

Midterm & Final Project Challenges

- **Project Format:**
  - Unlike the individual paper implementations, **these are challenge-based projects**
  - We will provide a **specific challenge topic and the necessary dataset**
  - Your task is to **use existing research papers as a foundation** to solve the challenge
- **Midterm Goal:**
  - **Implement** and **optimize** a relevant research paper to **complete the given image challenge task**.
- **Final Goal:**
  - Extend your midterm project
  - **Improve your results** by integrating ideas **from one or more additional papers**, or by **proposing your own novel approach**

Midterm & Final Challenge Topics

- Hair Density & Width Estimation **(2D Semantic & Instance Segmentation)**
- Scalp Analysis **(Image Classification, 2D Object Detection)**
- Surround-View RGB Object Detection **(2D Object Detection)**
- Lane Line Detection **(2D Semantic & Instance Segmentation)**
- Dual-Sensor (RGB + Thermal) Object Detection **(2D Object Detection)**
- High Dynamic Range (HDR) Imaging **(Image-to-Image, HDR Reconstruction)**
- Radar & RGB Fusion for Object Detection **(2D/3D Object Detection)**
- 3D LiDAR Object Detection **(3D object detection)**
- 3D LiDAR & RGB Fusion for Object Detection **(3D object detection)**
- 3D LiDAR SLAM **(Simultaneous Localization and Mapping)**

Important Dates & Deadlines

- **TA's Email:** D11102001@gapps.ntust.edu.tw
- **Regular Paper Reports:**
  - Oral Presentations (10 min): Paper 1 **(9/23)**, Paper3 **(11/11)**
  - Written Report & Code Submission: Paper 1 **(9/23)**, Paper 2 **(10/14)**, Paper3 **(11/11)**, Paper 4 **(12/2)**
- **Midterm Project:**
  - Topic Decision Deadline: **9/23** (Before Week 4)
  - Code Confirmation Period: **10/13** ~ **10/31** (**Email TA to schedule**)
  - Oral Presentation (20 min): **10/21**, **10/28**
  - Final Report & Code Submission: **10/31**
- **Final Project:**
  - Code Confirmation Period: **12/1** ~ **12/19** (**Email TA to schedule**)
  - Oral Presentation (20 min): **12/9**, **12/16**
  - Final Report & Code Submission: **12/19**

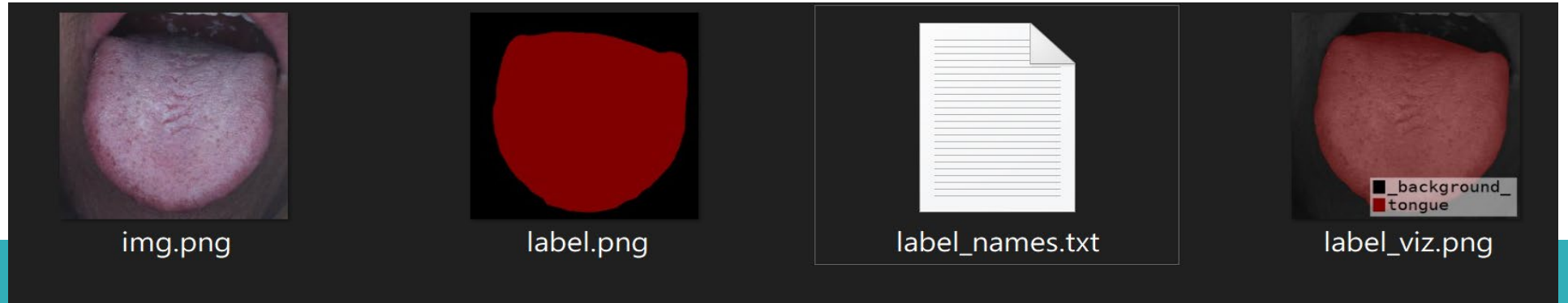
You must confirm your project topic with the TA. The **TA has the final say and may ask you to choose a different topic if it is unsuitable or selected by too many students**.

Homework 0: Skills & Hardware Verification

- **Objective:**
  - To verify your practical skills in **Python**, **ROS**, **Git**, **Docker**, **OpenCV**, and a **major deep learning framework** (PyTorch or TensorFlow).
- **Tasks:**
  - Task 1 (2D Segmentation): Perform tongue image segmentation using a 2D semantic segmentation model.
  - Task 2 (3D Detection): Train the DeepFusion network (a 3D object detection model using LiDAR & RGB data) and visualize the results on the KITTI dataset.
- **Grading & Deadline:**
  - Grading: This assignment is mandatory and accounts for 15% of your total participation grade.
  - Due: **Next Week (9/8~9/12)** (Please schedule a demo time with the TA).

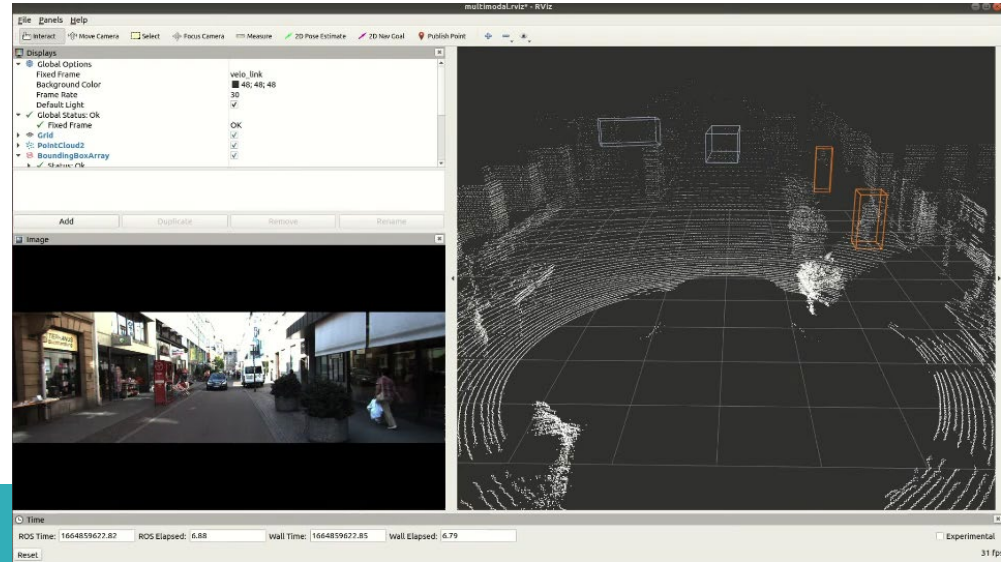
HW0 Task 1: Tongue Segmentation (2D Semantic Segmentation)

- **Task Description:**
  - Download the provided **Tongue Dataset** and train **a segmentation model** of your choice
  - Your implementation must use a **Docker** environment.
  - Upload your complete project to **GitHub**
  - Schedule a demo with the TA to test your model on a designated computer
- **Evaluation:**
  - Your model will be tested on a private dataset of 250 images.
  - Important Note: The TA will not provide any help or hints during the demo and testing process. You are on your own.
- **Links:** Tongue Dataset, MMSegmentation, Segmentation_models.pytorch



HW0 Task 2: LiDAR + RGB Object Detection (3D Object Detection)

- **Task Description:**
  - Download the KITTI Dataset and use it to train, validate, and visualize the DeepFusion model.
  - Your implementation must use a Docker environment.
  - Upload your complete project to GitHub
- **Deliverables:**
  - Your full project code on **GitHub**
  - **A 20-second demo video** showing your model's visualization results
- **Task Description:**
  - Email your GitHub link and the demo video to the TA
- **Links:** DeepFusion 、 KITTI Dataset



Thanks!

Any questions?

Sin-Ye Jhong (鍾昕燁)

E-mail: sinyejhong@mail.ntust.edu.tw