CSIE 2024 Fall - Computer Vision and Deep Learning (電腦視覺與深度學習)

Instructor: Jenn-Jier James Lien (連震杰), Professor, jjlien@mail.csie.ncku.edu.tw, Ext. 62540

Website: http://robotics.csie.ncku.edu.tw/course.html

Class Time/Location: 09:10 ~ 12:00 Friday / room A1302

Spoken Language: English **Prerequisites:** C, C++ or Python

Syllabus: (W: Week - Ch: Szeliski's book chapter)

	01-01	Course Introduction:
Computer Vision	01-02.1, 03.6	Sensor - **Camera model**: Geometric transformations between 2D and 3D.
	02-06.3	Sensor - **Camera calibration**: Optimization process and **AR** (Augmented reality).
	02-11.1, 11.3 12.1, 12.2	Sensor - **3D** (brief): 1) **Stereo**, 2) ToF (Time-Of-Flight, Kinect 2), and 3) Structured light (Kinect 1, DLP projector).
	03-04.1	**Detection/Extraction - SIFT** (and brief HOG): Feature extraction and descriptor.
	04-05.3.1, 12.6.4	**Motion Detection - Background subtraction/modeling**: Real-time motion detection using **GMM** (Gaussian mixture model).
	05-08.1	**Tracking - Optical flow**: Real-time feature point tracking using optical flow.
Assignment 1 Machine Learning	06-14.2, A.1	**Unsupervised Learning - PCA** (principal component analysis): Dimensionality reduction (domain knowledge) for face recognition (eigenface).
	07-14.2, A.1	**Supervised Learning for Classification - LDA** (linear discriminant analysis): Linear classification for face recognition.
Exam 1	08-	[From AI (artificial intelligence) to ML (machine learning), to DP (deep learning): -From Bayes' Rule (posterior probability) to Gaussian model, to similarity measure (likelihood probability: Mahalanobis distance, SSD (sum of squared differences) and correlation (or pattern matching)), to PCA (linear combination).]
	09-14.1	**Supervised Learning for Classification – Cascaded AdaBoost** (brief)**:** Face detection.
	09-14.1	**Supervised Learning for Classification - SVM** (support vector machines): Non-linear classification.
	10-03.7	**Unsupervised Learning for Classification - VQ** (vector quantization): Clustering and K-means.
	10-03.7	**HMM** (discrete-time hidden Markov model) (brief): Facial expression recognition in video.
Deep Learning	11-	**Classification** - Encoder (Backbone) and Head: LeNet and AlexNet
	12-	**Classification** - Encoder (Backbone) and Head: VGG16 and GoogLeNet
	12-	**Classification** - Encoder (Backbone) and Head: ResNet
Assignment 2	13-	**Segmentation and Classification** - Encoder (Backbone), Decoder (Neck) and Head: U-Net
	14-	**BBox Detection and Classification** - Encoder (Backbone), Decoder (Neck) and Head: Faster R-CNN and Mask R-CNN
Exam 2	15-	Introduction to **GPT**:
	16-	**GPT** (Generative Pre-trained Transformer) – **GAN** (Generative Adversarial Network) and **Pre-trained model**:
	17	**GPT** (Generative Pre-trained Transformer) – **Transformer model:**
	18-	**Final Project (Group Project) – Deep Learning**

Grading: Assignment x 2: 20% x 2=40%, Exam x 2: 20% x 2=20%, Project x 1: 20% x 1 = 20%.

Textbooks/Reference Books:

1. Textbook1: Class lecture notes.
2. Textbook2. Computer Vision: Algorithms and Applications by Richard Szeliski, Springer, V2, 2022.
3. Ref1: Learning OpenCV, Computer Vision with the OpenCV Library by Gary Bradski and Adrian Kaebler, O'Reilly, 2008.
4. Ref2. Deep Learning by Ian Goodfellow, Yoshua Bengio, and Aaron Courville, MIT Press, 2016.

課程概述：

這門課除了讓學生能深入了解電腦視覺、機器學習及人工智慧-深度學習的理論知識，與分析深度學習的原理是如何結合人工智慧類神經網路及電腦視覺發展而來的相互關係外，技術功能面會以授課老師 20 幾年的產學合作經驗（臉部偵測、辨識及表情分析、雲端智慧型監控服務、自動光學檢測、智慧製造、智慧型機器手臂控制、自走車）及近年來在臨床醫學影像與精準運動科技合作經驗來舉實際的範例解釋。

課程會先教攝影機及鏡頭產生的 2D 影像與 3D 物體及增擬實境之間關係的基本電腦視覺原理及技術開始，包括攝影機的校正方法及 3D 物體重建原理，接著就會傳授電腦視覺基本但實用的技術，包括即時偵測、追蹤及辨識系統的設計。再來藉由機器學習的連接帶入深度學習領域，教授如何藉由深度學習的原理來開發更好的即時偵測、追蹤及辨識技術以解決實際的產業界智慧製造自動化、臨床醫學影像及精準運動科技上的問題。

本課程期待培養學生於電腦視覺、機器學習及深度學習領域技術設計及整合實作的能力，透過務實的作業實例來培育學生具備研發思考、程式設計及解決現存問題的能力，藉由分組計畫的實作來培養學生具備發現問題、解決問題及團隊分工合作的能力與精神，並可把所學的理論基礎應用到工業界智慧製造自動化、臨床醫學影像處理及精準運動科技的實務面。

Course Abstract:

Apart from providing students with a deeper understanding of the theoretical knowledge of computer vision, machine learning and artificial intelligence - depth learning, and analyzing how the principles of depth learning combine with the development of artificial neural networks (ANNs) and computer vision, the functional aspects of this course will be explained by teacher's more than 20 years of experience in industrial cooperation (face detection, recognition and expression analysis, intelligent video surveillance as a service, automated optical inspection (AOI), intelligent manufacturing, visual-guided robot arm control, and automatic guided vehicle (AGV)).

The course will teach the basic computer vision principles and techniques for the relationship between 2D images and 3D objects with augmented reality (AR), including the calibration method of the camera and the reconstruction principle of 3D objects. Then I will teach the basic but practical techniques of computer vision, including the design of real-time detection, tracking and recognition systems. We will bring machine learning connections to the field of deep learning. Based on deep learning technology, I will teach you how to develop better real-time detection, tracking and

recognition techniques to solve practical problems in the fields of industry, clinical imaging and AI sports.

This course aims to develop students' ability to design and integrate skills in computer vision, machine learning and depth learning. Through practical practice, students' ability to think, program design and solve existing problems can be nurtured. Students' ability and spirit of team work can be nurtured through group work. They can also apply the theoretical foundation to industry, clinical imaging and precision sports analytics.

教學目標:

培養學生具備以下的基本知識及實作能力：

1）具備基本的電腦視覺、機器學習及深度學習理論基礎；

2）融匯貫通深度學習的原理是如何結合人工智慧及電腦視覺發展而來的；

3）具備以電腦視覺及深度學習的技術來解決即時偵測、追蹤、辨識及增擬實境在生活上、工業界（臉部偵測、辨識及表情分析、雲端智慧型監控服務、自動光學檢測、智慧製造、智慧型機器手臂控制、自走車）及近年來在臨床醫學影像與精準運動科技的實際問題。