

+886 936 503 850
franktpmvu@gmail.com

Yu-Hsi Chen



SUMMARY

Computer vision engineer with 9+ years of experience in deep learning and video/image processing, currently at Academia Sinica. Specialized in single and multi-object tracking, person re-identification, and robust recognition under adverse conditions. Recent projects integrate cutting-edge techniques such as latent diffusion models and YOLOv7 for synthetic-data-driven license plate recognition, achieving up to 95.05% recognition accuracy under adverse conditions such as rain and fog.

Applied Vision-Language Models for feature extraction to construct a Universal Action Space, and used LoRA fine-tuning to adapt models to downstream datasets while preserving the UAS structure. Actively exploring SAM2 and GroundingDINO to enable unsupervised video object segmentation and foreground-background detection for downstream visual understanding tasks.

WORK EXPERIENCE

Computer Vision Engineer (Research Assistant)

07 2015 - Present

Academia Sinica (Taiwan), Institute of Information Science(I.I.S.)

Taipei

- Designed and implemented deep learning models for computer vision tasks using Python and PyTorch, including CNNs, C3D, Siamese Networks, Transformers, and custom extensions of YOLOv7 for adverse conditions.
- NeighborTrack[Che+23]**: Achieved top performance on LaSOT and held #1 on UAV123 until 2024.
- Focus areas: object detection/tracking, person re-identification, video stabilization, and recent advances in VLMs, generative vision models, and foundation model applications.

EDUCATION

Master of Science in Electrical Engineering

09 2013 – 08 2015

Lunghwa University of Science and Technology

Taipei, Taiwan

Graduated **1st in department** (Rank 1/19, GPA: 92.34/100).

Master's thesis: *Full-Frame Video Stabilization via SIFT Feature Matching*.

Received the Excellent Paper Award at IIHMS 2014.

Bachelor of Science in Computer Information and Network Engineering

09 2009 – 07 2013

Lunghwa University of Science and Technology

Taipei, Taiwan

Focus areas: Verilog development for camera auto-exposure (AE) and auto-white balance (AWB); HDR imaging using MATLAB.

SELECTED PROJECTS

Video-Based Behavior Analysis via Long-Term Structured Modeling

2025

I.I.S. Research, Keywords: /S4/VST/V-JEPA/VideoCLIP/SAM2/GroundingDINO

- Developed a modular system for behavior analysis using vision-only, world, and vision-language models (VST, V-JEPA, VideoCLIP) with a Universal Action Space built from Kinetics-600.
- Applied structured state space models (S4/ViS4mer) for long-term temporal modeling, paired with a focal-loss-based key action selector to enhance focus on behavior-relevant segments.
- Investigated VLMs (SAM2, GroundingDINO) for unsupervised video object segmentation and foreground/background separation.

License Plate Recognition in Low Quality Images using Latent Diffusion [Che+24] 04 2024
I.I.S. Research, Keywords: YOLOv7/Latent Diffusion/Synthetic Data/Adverse Weather Simulation

- Created a synthetic license plate generation pipeline with simulated weather degradation, specifically targeting adverse conditions such as rain and fog to improve recognition robustness without relying on real-world training data.
- Developed LD-YOLOv7, an enhanced detection framework that integrates latent diffusion to restore noisy intermediate features and boost robustness in extreme weather.
- Achieved 87.38% accuracy on the AOLP dataset (on par with real-data-trained models) and 35.19% on the proposed WLP benchmark, outperforming prior methods in synthetic-to-real generalization.

NeighborTrack [Che+23] (Single Object Tracking) 06 2023
I.I.S. Research, keywords: Single Object Tracking/Context Modeling

- Developed NeighborTrack, a lightweight post-processing module that incorporates spatiotemporal context from nearby trajectories to mitigate identity switches in single object tracking.
 - Achieved 72.2% AUC on the LaSOT benchmark, surpassing competitive baselines.
 - Demonstrated broad compatibility with existing trackers without additional training.
- Project page: <https://github.com/franktpmvu/NeighborTrack>

[Early Work] Video Stabilization [CLS14] (Master's Thesis) 08 2014
keywords: SIFT/Content-Preserving Warping

- Developed a full-frame video stabilization system using SIFT feature matching and homography estimation for global motion extraction.
- Applied Kalman filtering and Gaussian smoothing to correct jitter while preserving intentional camera motion.
- Awarded Excellent Paper at IHHMSP 2014.

RESEARCH PUBLICATIONS

- [CLS14] Yu Hsi Chen, Hsueh Yi Sean Lin, and Chih Wen Su. “Full-Frame Video Stabilization via SIFT Feature Matching”. In: *2014 Tenth International Conference on Intelligent Information Hiding and Multimedia Signal Processing*. 2014, pp. 361–364. DOI: [10.1109/IIH-MSP.2014.96](https://doi.org/10.1109/IIH-MSP.2014.96).
- [Che+23] Yu-Hsi Chen et al. “NeighborTrack: Single Object Tracking by Bipartite Matching With Neighbor Tracklets and Its Applications to Sports”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*. June 2023, pp. 5138–5147.
- [Che+24] Yu-Hsi Chen et al. “License plate recognition in low quality image by using Latent Diffusion YOLOv7”. In: *2024 IEEE Conference on Artificial Intelligence (CAI)*. 2024, pp. 831–838. DOI: [10.1109/CAI59869.2024.00156](https://doi.org/10.1109/CAI59869.2024.00156).

SKILLS

Programming Python, MATLAB, L^AT_EX.

Development Tool Docker, Git.

Foundation Models Latent Diffusion, VideoCLIP, V-JEPA, SAM2, GroundingDINO.

Expertise Visual Modeling (object tracking, behavior recognition, detection), Temporal Modeling (S4, ViS4mer, Vision Transformers like VST), Video Understanding, Generative Vision (diffusion-based models).

Communication Chinese (native), English (basic professional use: reading papers), Japanese (intermediate, JLPT N3 certified).