

I-SHENG FANG

ishengfang.github.io

GitHub/Instagram \diamond ishengfang

+886 · 935 · 832173 \diamond nf0126@gmail.com \diamond ishengfang@citi.sinica.edu.tw

EDUCATION

Master of Science

September, 2020 - January, 2023

Robotics, National Yang Ming Chiao Tung University, the merger of National Chiao Tung University and National Yang Ming University.

Phi Tau Phi Scholastic Honor Society (Top 3 %)

Graduate student

January, 2020 - June, 2020

Computer Sciences, National Chengchi University.

Bachelor of Science

September, 2013 - January, 2018

Mathematical Sciences, National Chengchi University.

SELECTED PUBLICATIONS

Adapt2Hide: Leveraging Off-the-Shelf Autoencoder for Reversible Visual Processing

Ernie Chu*, *I-Sheng Fang**, Tai-Ming Huang, Pin-Yen Chiu, Vishal Patel, Jun-Cheng Chen

(* indicates equal contribution.)

IEEE International Conference on Image Processing (ICIP), 2026

Co-First author. Developing a novel image steganography approach using large pre-trained autoencoders (AEs) and low rank adaptation (LoRA), to achieve high-quality message reconstruction with minimal additional parameters and demonstrating effectiveness in practical reversible visual processing applications, such as reversible visual censorship, AI editing, and reversible style transfer.

Gen-n-Val: Agentic Image Data Generation and Validation

Jing-En Huang*, *I-Sheng Fang**, Tzuhsuan Huang*, Chih-Yu Wang, Jun-Cheng Chen

(* indicates equal contribution.)

IEEE/CVF Conference on Computer Vision and Pattern Recognition Findings Track (CVPRF), 2026.

[\[Paper\]](#)

Co-First author. Building agentic framework for image data generation and validation. Leveraging Layer Diffusion (LD), Large Language Model (LLM) as LD prompt agent, and Vision Large Language Model (VLLM) as data validation agent to generate data for improving performance of YOLOv9 and YOLO10 families, reducing invalid data rate from 50% to 7%, improving 1% mAP on rare classes in COCO instance segmentation, and 7.1% mAP in COCO open-vocabulary object detection benchmarks.

Text Slider: Efficient and Plug-and-Play Continuous Concept Control for Image/Video Synthesis via LoRA Adapters

Pin-Yen Chiu, *I-Sheng Fang*, Jun-Cheng Chen

IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), 2026.

Second Workshop on Visual Concepts (VisCon), IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2025.

[\[Paper \(arXiv, 8-Page Version\)\]](#) [\[Paper \(VisCon, OpenReview, 4-Page Version\)\]](#) [\[Project Page\]](#) [\[Code\]](#)

Proposed Text Slider, a lightweight and efficient method for concept control in diffusion models. Text Slider enables precise and continuous visual modulation by fine-tuning low-rank directions in the text encoder, achieving significantly better efficiency: 14 \times faster training than Concept Slider and 117 \times faster than Attribute Control, while also reducing GPU memory usage by 8 \times and 20 \times , respectively.

Every Camera Effect, Every Time, All at Once: 4D Gaussian Ray Tracing for Physics-based Camera Effect Data Generation

Yi-Ruei Liu*, You-Zhe Xie*, Yu-Hsiang Hsu*, *I-Sheng Fang*[†], Yu-Lun Liu, Jun-Cheng Chen
(* Equal contribution. Work done at Academia Sinica as intern. [†] Internship mentor.)

NeurIPS 2025 Workshop on SPACE in Vision, Language, and Embodied AI (SpaVLE). *Oral*.

[\[Paper \(arXiv\)\]](#) [\[Project Page\]](#)

Mentored interns. Developed 4D-GRT, a novel two-stage pipeline combining 4D Gaussian Splatting with physically based ray tracing to simulate real-world camera effects (e.g., fisheye distortion, rolling shutter, and depth of field). Achieved state-of-the-art rendering speed with competitive quality, and introduced a benchmark of eight synthetic dynamic scenes across four camera effects for evaluation.

CameraBench: Benchmarking Visual Reasoning in MLLMs via Photography

I-Sheng Fang, Jun-Cheng Chen

Multimodal Algorithmic Reasoning Workshop (MAR), IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2025.

[\[Paper \(arXiv\)\]](#) [\[Paper \(MAR\)\]](#)

Evaluating multimodal large language models (MLLMs) on photography-related visual reasoning tasks that require understanding how camera settings affect image appearance. Our benchmark reveals that current MLLMs struggle with consistent performance, highlighting the challenges and opportunities in developing models with deeper visual and physical understanding.

Camera Settings as Tokens: Modeling Photography on Latent Diffusion Models

I-Sheng Fang, Yue-Hua Han, Jun-Cheng Chen

SIGGRAPH Asia 2024 Conference Paper.

[\[Paper\]](#) [\[Project Page\]](#) [\[Demo\]](#) [\[Dataset\]](#) [\[Code\]](#) [\[Model\]](#)

Developing a novel approach for latent diffusion models (LDMs) by integrating camera settings into text-to-image models using a new curated dataset (CameraSettings20k) and embedding techniques (Camera Settings as Tokens) with LoRA adapter, significantly enhancing numerical control over generated images based on photographic principles. Our method also allow user to combine other method or LoRA adapter of LDMs, such as ControlNet, Dreambooth, and IP-Adapter.

Best of Both Sides: Integration of Absolute and Relative Depth Sensing Modalities Based on iToF and RGB Cameras

I-Sheng Fang, Wei-Chen Chiu, Yong-Sheng Chen

International Conference on Pattern Recognition (ICPR), 2024

[\[Paper\]](#) [\[Paper\]](#) [\[Code\]](#)

Integrating active sensing modality (indirect Time-of-Flight, iToF) and passive sensing modality (RGB) to estimate high-resolution metric depth without metric depth supervision. The model is weakly supervised by cross-warp (geometric) consistency and structure knowledge distillation.

ES³Net: Accurate and Efficient Edge-based Self-Supervised Stereo Matching Network

I-Sheng Fang, Hsuao-Chieh Wen, Chia-Lun Hsu, Po-Chung Jen, Ping-Yang Chen, Yong-Sheng Chen

Best Paper Award at The 19th Embedded Vision Workshop (EVW) which in conjunction with IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2023.

[\[Paper\(CVF\)\]](#) [\[Code\]](#)

Adapting efficient cost-volume-based stereo matching to self-supervised learning and deploying the network on the drone to improve the robustness. Achieved quasi-real-time depth estimation on NVIDIA Jetson TX2 and real-time depth estimation on NVIDIA Jetson AGX Orin. ES³Net only needs 4 minutes for fine-tuning with novel collected data.

Self-Contained Stylization via Steganography for Reverse and Serial Style Transfer

Hung-Yu Chen*, *I-Sheng Fang**, Chia-Ming Cheng, Wei-Chen Chiu.

(* indicates equal contribution.)

IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), 2020.

[\[Paper\(arxiv\)\]](#) [\[Paper\(CVF\)\]](#) [\[Project Page\]](#) [\[Code\]](#)

Co-First Author. Developed a two-stage model that integrates neural style transfer and deep steganography to address issues of reverse and serial style transfer. The model hides the features of the content image within the stylized image and later retrieves the encoded features of the content image for reconstructing original content image or further stylizing it.

OTHER PUBLICATIONS

ImagenWorld: Stress-Testing Image Generation Models with Explainable Human Evaluation on Open-ended Real-World Tasks

Samin Mahdizadeh Sani*, Max Ku*, Nima Jamali, Matina Mahdizadeh Sani, Paria Khoshtab, Wei-Chieh Sun, Parnian Fazel, Zhi Rui Tam, Thomas Chong, Edisy Kin Wai Chan, Donald Wai Tong Tsang, Chiao-Wei Hsu, Ting Wai Lam, Ho Yin Sam Ng, Chiafeng Chu, Chak-Wing Mak, Keming Wu, Hiu Tung Wong, Yik Chun Ho, Chi Ruan, Zhuofeng Li, *I-Sheng Fang*, Shih-Ying Yeh, Ho Kei Cheng, Ping Nie, Wenhui Chen

(* Equal contribution.)

The Fourteenth International Conference on Learning Representations (ICLR), 2026.

[\[Project Page\]](#) [\[Preprint\]](#) [\[Dataset\]](#) [\[Code\]](#) [\[Visualization\]](#) [\[Comfy UI Blog\]](#)

Contributed to the development of ImagenWorld, a large-scale benchmark designed to evaluate modern diffusion, autoregressive, and hybrid image-generation models across six core tasks and six diverse visual domains. Assisted in collecting and curating a dataset of 3.6K condition sets and 20K fine-grained human annotations, covering object-level and segment-level error labels for explainable evaluation.

KMOPS: Keypoint-Driven Method for Multi-Object Pose and Metric Size Estimation from Stereo Images

Ying-Kun Wu, Yi Shen, Tzuhsuan Huang, *I-Sheng Fang*, Jun-Cheng Chen

IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), 2026.

[\[Project Page\]](#) [\[Code\]](#)

Proposed a novel method (KMOPS) for estimating the 6-DoF pose and metric size of multiple objects from a single calibrated stereo image pair. The approach extracts and triangulates 2D keypoints from both views to recover metric 3D positions, enabling accurate pose and size estimation without predefined 3D search spaces or anchors. Demonstrated state-of-the-art performance on Transparent Object Dataset (TOD) and StereOBJ-1M.

Single Image Reflection Removal based on Knowledge-distilling Content Disentanglement

Yan-Tsung Peng, Kai-Han Cheng, *I-Sheng Fang*, Wen-Yi Peng, Jr-Shian Wu

IEEE Signal Processing Letters(SPL) Feb. 2022

[\[Paper\(IEEE Xplore\)\]](#) [\[Code\]](#)

Disentangling the reflection and transmission feature via knowledge distillation in order to remove the reflection with a single image.

PROJECTS

Justfont

July, 2021 - September, 2021; June, 2023 - September, 2023

AI Consultant

Implemented state-of-the-art generative models, integrated these models into Justfont's font design workflow, and provided strategic guidance, achieving a 15-30% reduction in design time for Justfont, a distinguished Taiwanese font company known for its typeface design and webfont service.

A Century of Heartfelt Sentiment:**100th Anniversary Special Exhibition of the Taiwan Cultural Association**

2021

The exhibition of National Museum of Taiwan Literature, Tainan, Taiwan.

Synthesizing DeepFake video for historical figure of Taiwan Cultural Association, an important organization during the Japanese rule of Taiwan.

Font Design with Progressive invariant GANs

2017

Honorable Mention at MOST Workshop on Generative Adversarial Networks and GAN Project Competition.

Introducing GLCM, a invariant, in progressive growing VAE-GAN to embed font in \mathbb{R}^{300} latent space. Generating new font or glyph by modified the latent space, improving converging time with higher resolution.

Hanzi Typeface Research with Conditional Generative Adversarial Network

2017

Poster of Taiwan Society for Industrial and Applied Mathematics Annual Meeting.

Using autoencoder and conditional GAN to embed font in \mathbb{R}^n latent space.

Advisor: Prof. Yen-Lung Tsai.

EXPERIENCE

Academia Sinica, Taiwan

March, 2024 -

Research Assistant

Working on the generative model at the Research Center for Information Technology Innovation (CITI) with Dr. Jun-Cheng Chen. Published one paper on the latent diffusion model and numerical camera settings at SIGGRAPH Asia 2024 as the first author. Presented one paper on benchmarking MLLMs via photography at MAR, CVPRW 2025 as the first author. Written two papers on generative models as the co-first author. Published one paper on generative models at WACV 2026 as the second author. Written one papers as the second author. Applied one US patent. Led two groups of summer intern in 2025. Written a proposal for Taiwan NSTC Research Project Grant.

Microsoft AI R&D Center, Taiwan

March, 2022 - November, 2022

Software Engineer Intern

Investigating vision transformer (ViT) as a backbone of perceptual loss and found that the task whether texture-driven or context-driven could affect the performance of the image-to-image translation model trained with ViT or ConvNet as the backbone of perceptual loss, working with SunDa Yang, Chien-Yi Wang, Prof. Shang-Hong Lai, Dr. Trista Chen, and the face science team.

National Chiao Tung University

September, 2018 - September, 2019

Research Assistant

Working on style transfer and generative model with Prof. Wei-Chen Chiu. Published the paper, Self-Contained Stylization via Steganography for Reverse and Serial Style Transfer, as co-first author at WACV 2020.

The First Step of AI and Deep Learning with Python

July, 2018 - September, 2018

Teaching Assistant

The MOOCs of National Chengchi University.

Writing the cue for video cut and additional tips and solving students' problems by online and offline. Teacher is Prof. Yen-lung Tsai.

InQtech Co.,Ltd

September, 2017 - June, 2018

Deep Learning Scientist

InQtech is AI and AIOT startup for deploying AI in manufacturing industry.
Investigating PCB segmentation by using semi-supervised GAN model to segment PCB for assembling.

The First Step of Data Analysis with Python

September, 2017 - January, 2018

Teaching Assistant

The MOOCs of National Chengchi University. Writing the cue for video cut and additional tips and solving students' problems by online and offline. Teacher is Prof. Yen-lung Tsai.

Leopard Mobile

July-September, 2017

intern

Data analysis in CM Security commerce realization team. Predicting user behavior prediction with deep learning model

NCCU Deep Learning & TWSIAM Student Group

January, 2017-January, 2018

Member

Winning **first** place of study group final presentation in NCCU. Hosting discussion about GAN.

SERVICES

Reviewer: NeurIPS 2024, CVPR 2025, ICCV 2025, NeurIPS 2025, NeurIPS SpaVLE 2025, WACV 2026, ICRA 2026, CVPR 2026, ICML 2026, ECCV 2026

Organizer: [Workshop on Generative AI for Photography, WACV 2026](#) (Primary Organizer)

Administrator: [Enjoyfonts \(a facebook group for typography\)](#)