

Yu-Teng Li (Kevin)

✉ yutengli@berkeley.edu  linkedin.com/in/yutengli  415-418-0756  thekevinli.github.io

EDUCATION

University of California, Berkeley
B.S. Electrical Engineering & Computer Science

Aug 2019 – May 2023
(GPA: 3.7 / 4)

SELECTED PUBLICATIONS

UniFusion: VLM as Unified Encoder in Image Generation and Editing [[ArXiv](#)] [[Website](#)] – ICLR 2026 Multimodal Intelligence Workshop. UniFusion challenges the premise that semantic embeddings are insufficient for pixel generation tasks. Shows that multi-layer VLM features alone enable high-quality T2I generation and editing diffusion models (rivaling Flux and Qwen-Image-Edit), without VAE conditioning. The unified multimodal space further yields emergent capabilities such as zero-shot multi-reference generation and cross-task knowledge transfer.

Towards Text-Guided Attribute-Disentangled Multimodal Representation Learning – CVPR 2026. Formulates Queryable Attribute Representation Extraction (QARE) and introduces the first benchmark for attribute-disentangled multimodal embeddings. Proposes a training-free method that substantially outperforms SigLIP and contrastively-trained methods such as VLM2Vec, on attribute-conditioned retrieval.

Hyperbolic Active Learning for Semantic Segmentation under Domain Shift [[ArXiv](#)] – ICML 2024. HALO uses a hyperbolic neural network for pixel-level active learning in semantic segmentation, and is the first active learning approach that surpasses the performance of supervised domain adaptation with only 1% labeled data on synthetic-to-real benchmarks such as GTAV → Cityscapes.

Neighboring State-based Exploration for Reinforcement Learning [[ArXiv](#)] – Inspired by adversarial attack literature, we propose a simple but effective on-policy exploration method by surveying a bounded region of nearby states during early training of an agent. Our method consistently outperforms Double DQN baseline by 49% in discrete environments. Built in *PyTorch*, *MuJoCo*.

Squeezing Capacity from Multimodal LLMs for Subject-driven Generation – In submission to NeurIPS 2026. Achieved superior identity preservation and harmonization of subjects than all previous arts by aggregating multi-layer VLM features and a multi-stage denoising strategy to balance the details in VAE and VLM cross-modal semantics.

EXPERIENCE

Senior Applied Research Scientist, Firefly, [Adobe Inc.](#), San Jose, CA.

Aug 2023 – Present

- Led multi-reference editing for next-generation Firefly **video & image generalist Mixture-of-Experts (MoE)** model. Ran training ablations on cross-modal editing recipes and VLM conditioning. Built activation analysis tools that guided improvements in expert routing.
- Co-led multimodal pretraining for Firefly Image 5**, a production model supporting 4MP text-to-image, instruction editing, character reference and layered generation. Designed data ablation studies and drove the final training recipe to scale on 1000+ GPUs with billion-scale datasets.
- Led aesthetic finetuning (SFT) for **Firefly Image 4** foundation model, delivering the recipe demoed at Adobe MAX. Introduced a timestep curriculum during pretraining assigning GenAI samples to high-noise diffusion steps for broader concept coverage without saturation bias.
- Led personalization for **Firefly Image 3** – improved DreamBooth recipe's stability with VLM-predicted concept and super-class, improved memory efficiency with AdaFactor, enabled ControlNet and Stylization, leading to multi-million-dollar revenues on enterprise customers.
- Proposed **UniFusion**, a Diffusion Transformer architecture replacing separate text and image encoders, with multi-layer VLM semantic features, which achieves competitive generation & editing, and **zero-shot multi-image reference generation**.
- Proposed **inversion-based style transfer with weighted cross-attention**, a training-free approach for synthesizing high-quality stylization pairs for instruction-editing finetuning.

Student Researcher, [Berkeley Artificial Intelligence Research](#), Berkeley, CA.

Feb 2022 – Oct 2023

- Led research projects on **Active Learning on Image Segmentation** for autonomous driving, in Trevor Darrell's group. [[lab website](#)]
- Developed a config-driven experiment package which supports image-, pixel-, and region-based active learning (AL) with multiple sampling heuristics (e.g. entropy, mask autoencoding), eventually leading to the creation of **HALO, the first AL method to surpass fully supervised baseline** in segmentation benchmarks (ICML 2024).

Software Engineering Intern, [Adobe Inc.](#), San Jose, CA.

May 2022 – Aug 2022

- Designed and trained a **CycleGAN**-based facial expression translation to **mitigate warping artifacts** in Photoshop **Liquify Filter**.
- Developed **Smart Grid for Face Stylization** – a real-time web app (*10 fps*) that stylizes portraits via GAN latent space interpolation. Designed a **layout algorithm** organizing 10k StyleGAN2 embeddings into grids using UMAP and K-means on style affinity scores.

Student Instructor (Head TA of Discussion), CS182 Deep Learning, [UC Berkeley](#).

Aug 2022 – May 2023

- Led the discussion curriculum design and taught sections on VAE, diffusion models and transformers (300+ students, Spring 2023).

Student Researcher, [Vision and Learning Lab @ National Taiwan University](#).

Jan 2021 – Oct 2021

Machine Learning Intern, [Dell Technologies Inc.](#), Taiwan.

Jul 2021 – Aug 2021