

TRUONG VU

vu.tru@northeastern.edu | [linkedin.com/in/vu-tru/](https://www.linkedin.com/in/vu-tru/) | scholar.google.com/citations?user=mTbQ2ZsAAAAJ

EDUCATION

Northeastern University

Ph.D. Computer Science

Advisor: Dr. Ehsan Elhamifar

Research Interests: Generative Models, Diffusion models, Computer Vision, Egocentric Vision, Video Generation.

Boston, MA

Aug. 2024 - Present

Hanoi University of Science and Technology

Bachelor of Engineering, Computer Science

Advisor: Dr. Khoat Than

Research Interests: Computer Vision, Generative Models, Generative Adversarial Networks.

Hanoi, Vietnam

Sep. 2018 – July 2022

PUBLICATIONS

- **Truong Vu**, Daiwei Zhang, Ehsan Elhamifar, "EgoAcGen: Interaction-Aware Egocentric Action Generation", (**Submission ECCV 2026**)
- Ho Quang Nguyen*, **Truong Vu***, Anh Tran, Khoi Nguyen, "Dataset diffusion: Diffusion-based synthetic data generation for pixel-level semantic segmentation", (**NeurIPS 2024**) [PDF] [Project Page]
- Chau Pham*, **Truong Vu***, and Khoi Nguyen, "LP-OVOD: Open-vocabulary object detection by linear probing", (**WACV 2024**) [PDF] [Code]
- **Truong Vu**, Kien Do, Khang Nguyen, and Khoat Than, "Face swapping as a simple arithmetic operation", [PDF] [Code]
- Khang Nguyen, Kien Do, **Truong Vu**, and Khoat Than, "Unsupervised image segmentation with robust virtual class contrast", (**Pattern Recognition Letters 2023**) [PDF]

WORK EXPERIENCE

Graduate Research Assistant

NOVA Lab, Northeastern University

Sept. 2024 – Present

Boston, MA

Project: Egocentric Reference Video Generation (ongoing work)

- Developing a generative video model that transforms **third-person** instructional videos into **first-person** egocentric reference videos for performing the same action.
- Aiming to support embodied AI assistants that can provide personalized visual guidance for tasks such as cooking, assembly, and daily activities.

Project: Interaction-Aware Egocentric Action Generation

- *Problem*: Recent video diffusion models struggle with egocentric views: they often render hands unrealistically, create inconsistent hand-object interactions.
- Developed a novel framework to enhance video generation models so that they encode 3D hand pose and 2D hand-object information. The method yields sharp and temporally consistent hand-object interactions with **~ 10% improvement** on hand-object consistency.

AI Resident

VinAI Research (Now Qualcomm AI Research)

July. 2022 – July 2024

Hanoi, Vietnam

Project: Diffusion-based Synthetic Dataset Generation for Pixel-Level Semantic Segmentation.

- *Problem*: Annotating large-scale datasets with pixel-level labels is labor-intensive and time-consuming.
- Developed an **attention-based** mask refinement and **uncertainty-aware** training framework to generate paired synthetic images and segmentation masks from Stable Diffusion *without human annotation*.
- Synthetic-only training achieved **~ 81%** of fully supervised real-data performance.

Project: Open-Vocabulary Object Detection.

- *Problem*: CLIP-based OVOD methods often assign high scores to low-quality box proposals because CLIP is not trained for precise object localization.
- Proposed **LP-OVOD**, a linear-probing approach that filters low-quality proposals using pseudo labels retrieved from CLIP-based region-text matching.
- Improved novel-class detection performance significantly by **~ 44%** without using external datasets or requiring novel-class annotations during training.

Research Assistant

Data Science Laboratory, Hanoi University of Science and Technology.

July 2020 – May, 2022

Hanoi, Vietnam

Project: Face Swapping with StyleGAN.

- *Problem:* Face swapping is the task of transferring the identity from a source face image to a target face image while preserving the attributes (styles) of the target face image.
- Proposed a novel high-fidelity face swapping method called Arithmetic Face Swapping (AFS) that explicitly disentangles the intermediate latent space $W+$ of a pretrained StyleGAN.

Project: Unsupervised image segmentation.

- *Problem:* Unsupervised image segmentation (UIS), which aimed to classify pixels in an image into different semantic classes without using any labels.
- Proposed a novel method for unsupervised image segmentation called Robust Virtual Class Contrast (RVCC), which characterized foreground pixels based on their virtual classes rather than just saliency masks.

ENGINEERING SKILLS

- **Programming Languages:** Python, C/C++, Java, SQL, Javascript
- **ML Frameworks/Tools:** Pytorch, Keras, Tensorflow, Jupyter, OpenCV, Numpy, Scipy
- **Others:** Git, AWS, Slurm, Docker, Flask, Django, Google Cloud Platform, Apptainer

SERVICES

- Served as a reviewer for CVPR, ICCV, and NeurIPS conferences.

HONORS AND AWARDS

- Gold Medals, National Math Competition for University Students, 2019
- Silver Medals, International Mathematics Tournament of Towns (ITOT), 2017
- Best Presentation Thesis Award, Hanoi University of Science and Technology, 2022
- Academic Encouragement Scholarship, Hanoi University of Science and Technology, 2020