

Yuan-Hong Liao

✉ andrew@cs.toronto.edu | 🌐 <https://andrewliao11.github.io> | 📄 [andrewliao11](#) | 📺 [andrewliao11](#) | 🐦 [@andrewliao11](#)

Research Interests

Multi-modal LLMs, Efficient Data Annotation, and Dataset Improvement

Education

University of Toronto

PH.D. IN COMPUTER SCIENCE
• Advisor: Prof. Sanja Fidler

Ontario, Canada
Sep. 2019 - Jun. 2025 (expected)

National Tsing-Hua University

B.S. IN ELECTRICAL ENGINEERING
• Advisor: Prof. Min Sun, Rank: top 10%

Hsinchu, Taiwan
Sep. 2013 - Jun. 2017

Work Experience

Amazon Applied Scientist Intern

AT AMAZON CoRo TEAMS, I WORK ON FINE-TUNING LARGE VISION-LANGUAGE MODELS FOR IN-CONTEXT LEARNING.

Manager: Dr. Albert Chen
2024 Jul. - 2024 Oct.

Nvidia Research Scientist Intern

WE PUBLISH TWO RESEACH PAPERS *CARE* IN TMLR'23 AND *label transfer* IN ICLR'24. WE ORGANIZED AN INTERNAL DATA-CENTRIC SERIES AND I GAVE A TALK REGARDING DATA ANNOTATION AND LABEL AGGREGATION.

Manager: Dr. Sanja Fidler
2022 Jun. - 2023 May

Selected Publication

Can Feedback Enhance Semantic Grounding in Large Vision-Language Models?

WE INVESTIGATE HOW VISION-LANAUGES MODELS (VLMs) CAN IMPROVE BY TAKING FEEDBACK. WITHOUT ANY ADDITIONAL DATA, MODEL ARCHITECTURAL CHANGES, OR TRAINING, WE FIND THAT PROMPT-BASED SELF-FEEDBACK IMPROVES SEMANTIC GROUNDING IN VLMs BY UP TO 8 ACCURACY POINTS. [PAPER](#)
[Yuan-Hong Liao](#), Rafid Mahmood, Sanja Fidler, David Acuna

In submission

Reasoning Paths with Reference Objects Elicit Quantitative Spatial Reasoning in Large Vision-Language Models

WE INTRODUCE Q-SPATIAL BENCH, A QUANTITATIVE SPATIAL REASONING BENCHMARK, AND PROPOSE SPATIALPROMPT THAT ELICITS QUANTITATIVE SPATIAL REASONING. [PAPER](#)
[Yuan-Hong Liao](#), Rafid Mahmood, Sanja Fidler, David Acuna

EMNLP2024 - Main

Translating Labels to Solve Annotation Mismatches Across Object Detection Datasets

WE IDENTIFY A COMMON BUT UNDER-EXPLORED LABEL ISSUES BETWEEN DATASETS – ANNOTATION MISMATCHES. WE PROPOSE AN DATA-DRIVEN & DATA-CENTRIC APPROACH TO MITIGATE ANNOTATION MISMATCHES, IMPROVING DOWNSTREAM OBJECT DETECTORS ACROSS SEVEN DATASETS AND THREE OBJECT DETECTORS. [PAPER](#)
[Yuan-Hong Liao](#), David Acuna, Rafid Mahmood, James Lucas, Viraj Prabhu, Sanja Fidler

ICLR 2024

Bridging the Sim2Real gap with CARE: Supervised Detection Adaptation with Conditional Alignment and Reweighting

WE STUDY THE SETTING OF SUPERVISED SIM2REAL DA APPLIED TO 2D OBJECT DETECTION. OUR ALGORITHM SYSTEMATICALLY EXPLOITS TARGET LABELS TO CLOSE THE SIM2REAL APPEARANCE AND CONTENT GAPS. [PAPER](#)
Viraj Prabhu, David Acuna, Rafid Mahmood, Marc T. Law, Yuan-Hong Liao, Judy Hoffman, Sanja Fidler, James Lucas

TMLR 2023

LA-BALD: An Information-Theoretic Image Labeling Task Sampler

WE PROPOSE A NOVEL ACTIVE SAMPLER, LA-BALD, THAT EXPLICITLY CONSIDERS THE DOWNSTREAM LABEL AGGREGATION.
[Yuan-Hong Liao](#), Sanja Fidler

arXiv 2022

Towards Good Practices for Efficiently Annotating Large-Scale Image Classification Datasets

Oral presentation, CVPR 2021

WE EXTENSIVELY ANALYZE VARIOUS PRACTICES OF INCORPORATING MACHINE MODELS INSIDE THE ONLINE HUMAN ANNOTATIONS. [PROJECT WEBSITE](#), [DEMO](#)

[Yuan-Hong Liao](#), [Amlan Kar](#), [Sanja Fidler](#)

Emergent Road Rules In Multi-Agent Driving Environments

ICLR 2021

WE ANALYZE WHAT INGREDIENTS IN DRIVING ENVIRONMENTS CAUSE THE EMERGENCE OF THESE ROAD RULES AND FIND THAT TWO CRUCIAL FACTORS ARE NOISY PERCEPTION AND AGENTS' SPATIAL DENSITY. [PROJECT WEBSITE](#)

[Avik Pal](#), [Jonah Philion](#), [Yuan-Hong Liao](#), [Sanja Fidler](#)

Watch-And-Help: A Challenge for Social Perception and Human-AI Collaboration

ICLR 2021

WE BUILD VIRTUALHOME-SOCIAL, A MULTI-AGENT HOUSEHOLD ENVIRONMENT, AND PROVIDE A BENCHMARK INCLUDING BOTH PLANNING AND LEARNING BASED BASELINES. [PROJECT WEBSITE](#)

[Xavier Puig](#), [Tianmin Shu](#), [Shuang Li](#), [Zilin Wang](#), [Yuan-Hong Liao](#), [Joshua B. Tenenbaum](#), [Sanja Fidler](#), [Antonio Torralba](#)

Synthesizing Environment-Aware Activities via Activity Sketches

CVPR 2019

WE PRODUCE PROGRAMS AFFORDED BY THE GIVEN ENVIRONMENT (ENVIRONMENT-AWARE PROGRAM GENERATION).

WE CREATE VIRTUALHOME-ENV AS OUR TESTBED FOR THE PROPOSED RESACTGRAPH. [PROJECT WEBSITE](#)

[Yuan-Hong Liao*](#), [Xavier Puig*](#), [Marko Boben](#), [Antonio Torralba](#), [Sanja Fidler](#)

Show, Adapt, and Tell of Cross-Domain Image Captioner

ICCV 2017

WE PROPOSE TO USE TWO CRITICS, DOMAIN CRITIC AND MULTI-MODAL CRITIC, TO GUIDE THE CAPTIONER TO PERFORM DOMAIN ADAPTATION.

[Tseng-Hung Chen](#), [Yuan-Hong Liao](#), [Ching-Yao Chuang](#), [Wan-Ting Hsu](#), [Jianlong Fu](#), [Min Sun](#)

Teaching

TA in Neural Networks and Deep Learning

CSC413/2516 Winter 2020

UNIVERSITY OF TORONTO

TA in The Cutting-Edge of Deep Learning

Fall 2017

NATIONAL TSING-HUA UNIVERSITY

TA in Signal and System

Spring 2017

NATIONAL TSING-HUA UNIVERSITY