Tsu-Jui (Ray) Fu

tsujuifu@gmail.com tsujuifu.github.io linkedin.com/in/tsujuifu1996

I am a research scientist at Apple AI/ML. My research lies in vision+language and text-guided visual editing. I obtained my Ph.D. in computer science from UC Santa Barbara, advised by William Yang Wang. I am also interested in video summarization and information extraction. My goal is to bridge the gap between multiple modalities via the AI system.

Education

Ph.D. in Computer Science, *UC Santa Barbara* 2019 - 2024
B.S. in Computer Science, *National Tsing Hua University* 2014 - 2018

Work Experience

Apple AI/ML
Research Scientist 2024 - Present
- Building large-scale multimodal foundation models

UC Santa Barbara
Research Assistant, advised by William Yang Wang 2019 - 2024
- Research in vision+language perception, exploration, and manipulation

Academia Sinica
Research Assistant, advised by Wei-Yun Ma 2018 - 2019
- Research in machine comprehension and information extraction

Publication (* indicates equal contribution)

TC-Bench: Benchmarking Temporal Compositionality in Text-to-Video and Image-to-Video Generation
Weixi Feng, Jiachen Li, Michael Saxon, **Tsu-jui Fu**, Wenhu Chen, and William Yang Wang
arXiv:2406.08656, 2024

T2V-Turbo: Breaking the Quality Bottleneck of Video Consistency Model with Mixed Reward Feedback
Jiachen Li, Weixi Feng, **Tsu-Jui Fu**, Xinyi Wang, Sugato Basu, Wenhu Chen, and William Yang Wang
arXiv:2405.18750, 2024

From Text to Pixel: Advancing Long-Context Understanding in MLLMs
Yujie Lu, Xiujun Li, **Tsu-Jui Fu**, Miguel Eckstein, and William Yang Wang
arXiv:2405.14213, 2024

Ferret-v2: An Improved Baseline for Referring and Grounding with Large Language Models
Haotian Zhang*, Haoxuan You*, Philipp Dufter, Bowen Zhang, Chen Chen, Hong-You Chen, **Tsu-Jui Fu**, William Yang Wang, Shih-Fu Chang, Zhe Gan, and Yinfei Yang
Conference on Language Modeling (**COLM**), 2024

Guiding Instruction-based Image Editing via Multimodal Large Language Models
Tsu-Jui Fu, Wenze Hu, Xianzhi Du, William Yang Wang, Yinfei Yang, and Zhe Gan
International Conference on Learning Representations (**ICLR**), 2024 (**Spotlight**)

VELMA: Verbalization Embodiment of LLM Agents for Vision and Language Navigation in Street View
Raphael Schumann, Wanrong Zhu, Weixi Feng, **Tsu-Jui Fu**, Stefan Riezler, and William Yang Wang
Association for the Advancement of Artificial Intelligence (**AAAI**), 2024

Discriminative Diffusion Models as Few-shot Vision and Language Learners
Xuehai He, Weixi Feng, **Tsu-Jui Fu**, Varun Jampani, Arjun Akula, Pradyumna Narayana, Sugato Basu, William Yang Wang, and Xin Eric Wang
arXiv:2305.10722, 2023

Text-guided 3D Human Generation from 2D Collections
Tsu-Jui Fu, Wenhan Xiong, Yixin Nie, Jingyu Liu, Barlas Oguz, and William Yang Wang
Conference on Empirical Methods in Natural Language Processing (**EMNLP**), 2023 (Findings)

EDIS: Entity-Driven Image Search over Multimodal Web Content
Siqi Liu*, Weixi Feng*, **Tsu-Jui Fu**, Wenhu Chen, and William Yang Wang
Conference on Empirical Methods in Natural Language Processing (**EMNLP**), 2023 (Long)

Collaborative Generative AI: Integrating GPT-k for Efficient Editing in Text-to-Image Generation
Wanrong Zhu, Xinyi Wang, Yujie Lu, **Tsu-Jui Fu**, Xin Eric Wang, Miguel Eckstein, and William Yang Wang
Conference on Empirical Methods in Natural Language Processing (**EMNLP**), 2023 (Short)

Photoswap: Personalized Subject Swapping in Images
Jing Gu, Yilin Wang, Nanxuan Zhao, **Tsu-Jui Fu**, Wei Xiong, Qing Liu, Zhifei Zhang, He Zhang, Jianming Zhang, HyunJoon Jung, and Xin Eric Wang
Conference on Neural Information Processing Systems (**NeurIPS**), 2023

LayoutGPT: Compositional Visual Planning and Generation with Large Language Models
Weixi Feng*, Wanrong Zhu*, **Tsu-Jui Fu**, Varun Jampani, Arjun Akula, Xuehai He, Sugato Basu, Xin Eric Wang, and William Yang Wang
Conference on Neural Information Processing Systems (**NeurIPS**), 2023

Tell Me What Happened: Unifying Text-guided Video Completion via Multimodal Masked Video Generation
Tsu-Jui Fu, Licheng Yu, Ning Zhang, Cheng-Yang Fu, Jong-Chyi Su, William Yang Wang, and Sean Bell
Conference on Computer Vision and Pattern Recognition (**CVPR**), 2023

An Empirical Study of End-to-End Video-Language Transformers with Masked Visual Modeling
Tsu-Jui Fu*, Linjie Li*, Zhe Gan, Kevin Lin, William Yang Wang, Lijuan Wang, and Zicheng Liu
Conference on Computer Vision and Pattern Recognition (**CVPR**), 2023

Training-Free Structured Diffusion Guidance for Compositional Text-to-Image Synthesis
Weixi Feng, Xuehai He, **Tsu-Jui Fu**, Varun Jampani, Arjun Akula, Pradyumna Narayana, Sugato Basu, Xin Eric Wang, and William Yang Wang
International Conference on Learning Representations (**ICLR**), 2023

ULN: Towards Underspecified Vision-and-Language Navigation
Weixi Feng, **Tsu-Jui Fu**, Yujie Lu, and William Yang Wang
Conference on Empirical Methods in Natural Language Processing (**EMNLP**), 2022 (Long)

CPL: Counterfactual Prompt Learning for Vision and Language Models
Xuehai He, Diji Yang, Weixi Feng, **Tsu-Jui Fu**, Arjun Akula, Varun Jampani, Pradyumna Narayana, Sugato Basu, William Yang Wang, and Xin Eric Wang
Conference on Empirical Methods in Natural Language Processing (**EMNLP**), 2022 (Long)

Language-Driven Artistic Style Transfer
Tsu-Jui Fu, Xin Eric Wang, and William Yang Wang
European Conference on Computer Vision (**ECCV**), 2022

M^{3}L: Language-based Video Editing via Multi-Modal Multi-Level Transformers
Tsu-Jui Fu, Xin Eric Wang, Scott Grafton, Miguel Eckstein, and William Yang Wang
Conference on Computer Vision and Pattern Recognition (**CVPR**), 2022

DOC2PPT: Automatic Presentation Slides Generation from Scientific Documents
Tsu-Jui Fu, William Yang Wang, Daniel McDuff, and Yale Song
Association for the Advancement of Artificial Intelligence (**AAAI**), 2022

VIOLET: End-to-End Video-Language Transformers with Masked Visual-token Modeling
Tsu-Jui Fu, Linjie Li, Zhe Gan, Kevin Lin, William Yang Wang, Lijuan Wang, and Zicheng Liu
arXiv:2111.12681, 2021

H-FND: Hierarchical False-Negative Denoising for Robust Distantly-Supervised Relation Extraction
Jhih-Wei Chen*, **Tsu-Jui Fu***, Chen-Kang Lee, and Wei-Yun Ma
Annual Meeting of the Association for Computational Linguistics (**ACL**), 2021 (Findings)

Semi-Supervised Policy Initialization for Playing Games with Language Hints
Tsu-Jui Fu and William Yang Wang
North American Chapter of the Association for Computational Linguistics (**NAACL**), 2021 (Short)

L2C: Describing Visual Differences Needs Semantic Understanding of Individuals
An Yang, Xin Wang, **Tsu-Jui Fu**, and William Yang Wang
European Chapter of the Association for Computational Linguistics (**EACL**), 2021 (Short)

Multimodal Style Transfer Learning for Outdoor Vision-and-Language Navigation
Wanrong Zhu, Xin Wang, **Tsu-Jui Fu**, An Yan, Pradyumna Narayana, Kazoo Sone, Sugato Basu, and William Yang Wang
European Chapter of the Association for Computational Linguistics (**EACL**), 2021 (Long)

SSCR: Iterative Language-Based Image Editing via Self-Supervised Counterfactual Reasoning
Tsu-Jui Fu, Xin Eric Wang, Scott Grafton, Miguel Eckstein, and William Yang Wang
Conference on Empirical Methods in Natural Language Processing (**EMNLP**), 2020 (**Oral**)

Counterfactual Vision-and-Language Navigation via Adversarial Path Sampler
Tsu-Jui Fu, Xin Eric Wang, Matthew Peterson, Scott Grafton, Miguel Eckstein, and William Yang Wang
European Conference on Computer Vision (**ECCV**), 2020 (**Spotlight**)

Why Attention? Analyzing and Remedying BiLSTM Deficiency in Modeling Cross-Context for NER
Peng-Hsuan Li, **Tsu-Jui Fu**, and Wei-Yun Ma
Association for the Advancement of Artificial Intelligence (**AAAI**), 2020 (**Oral**)

Learning from Observation-Only Demonstration for Task-Oriented Language Grounding via Self-Examination
Tsu-Jui Fu, Yuta Tsuboi, Sosuke Kobayashi, and Yuta Kikuchi
Conference on Neural Information Processing Systems (**NeurIPS**), 2019 (ViGIL workshop)

A Distributed Scheme for Accelerating Semantic Video Segmentation on An Embedded Cluster
Hsuan-Kung Yang*, **Tsu-Jui Fu***, Kuan-Wei Ho, Po-Han Chiang and Chun-Yi Lee
International Conference on Computer Design (**ICCD**), 2019 (**Oral**)

Adversarial Active Exploration for Inverse Dynamics Model Learning
Zhang-Wei Hong, **Tsu-Jui Fu**, Tzu-Yun Shann, Yi-Hsiang Chang, and Chun-Yi Lee
Conference on Robot Learning (**CoRL**), 2019 (**Oral**)

GraphRel: Modeling Text as Relational Graphs for Joint Entity and Relation Extraction
Tsu-Jui Fu, Peng-Hsuan Li, and Wei-Yun Ma
Annual Meeting of the Association for Computational Linguistics (**ACL**), 2019 (Long)

Attentive and Adversarial Learning for Video Summarization
Tsu-Jui Fu, Shao-Heng Tai, and Hwann-Tzong Chen
Winter Conference on Applications of Computer Vision (**WACV**), 2019 (**Oral**)

Region-Semantics Preserving Image Synthesis
Kang-Jun Liu, **Tsu-Jui Fu**, and Shan-Hung Wu
Asian Conference on Computer Vision (**ACCV**), 2018

Diversity-Driven Exploration Strategy for Deep Reinforcement Learning
Zhang-Wei Hong, Tzu-Yun Shann, Shih-Yang Su, Yi-Hsiang Chang, **Tsu-Jui Fu**, and Chun-Yi Lee
Conference on Neural Information Processing Systems (**NeurIPS**), 2018

Speed Reading: Learning to Read ForBackward via Shuttle
Tsu-Jui Fu and Wei-Yun Ma
Conference on Empirical Methods in Natural Language Processing (**EMNLP**), 2018 (Long)

Visual Relationship Prediction via Label Clustering and Incorporation of Depth Information
Hsuan-Kung Yang, An-Chieh Cheng*, Kuan-Wei Ho*, **Tsu-Jui Fu**, and Chun-Yi Lee
European Conference on Computer Vision (**ECCV**), 2018 (PIC workshop)

Dynamic Video Segmentation Network
Yu-Syuan Xu, **Tsu-Jui Fu***, Hsuan-Kung Yang*, and Chun-Yi Lee
Conference on Computer Vision and Pattern Recognition (**CVPR**), 2018

Intern Experience

Apple AI/ML
Research Intern, advised by Zhe Gan and Yinfei Yang — Summer 2023
- Leveraged multimodal large language models to improve instruction-based image editing

Meta AI
Research Intern, advised by Licheng Yu and Sean Bell — Summer 2022
- Developed multimodal masked video generation to unify text-guided video completion

Microsoft Azure AI
Research Intern, advised by Linjie Li, Zhe Gan, and Lijuan Wang — Summer 2021
- Developed masked visual modeling to improve large-scale text-video pre-training

Microsoft Research
Research Intern, advised by Yale Song and Daniel McDuff — Summer 2020
- Developed automatic document-to-slide system

Preferred Networks
Research Intern, advised by Yuta Tsuboi and Jason Naradowsky — Summer 2019
- Developed self-examination in imitation learning to improve task-oriented language grounding

Organizer

3rd Workshop on Advances in Language and Vision Research (ALVR)
Annual Meeting of the Association for Computational Linguistics (**ACL**), 2024

2nd Workshop on Advances in Language and Vision Research (ALVR)
North American Chapter of the Association for Computational Linguistics (**NAACL**), 2021

Award

Outstanding Dissertation Award, *UC Santa Barbara* — Spring 2024
Second Place, Alexa Prize SimBot Challenge, *Amazon* — Spring 2023
Third Place, Formosa Grand Challenge, *Taiwan* — Spring 2019
Second Place, Person In Context Challenge, *ECCV'18* — Summer 2018

Teaching

CS190I Introduction to Natural Language Processing, *UC Santa Barbara* — Winter 2022
CS565600 Deep Learning, *National Tsing Hua University* — Fall 2017
CS210401 Hardware Design and Lab, *National Tsing Hua University* — Fall 2017