JIALI DUAN

jli.duan@gmail.com ◇ Linkedin ◇ Website ◇ USC Alumni

SUMMARY

I'm a senior Research Scientist working on Apple's Foundational Diffusion Models. My current focus is building video & world generation models, and post-training agents that use them as tools to align generations with human preference. I'm also the core contributor to **Image Playground** and **Genmoji** in Apple Intelligence.

Previously, I was a research scientist at FAIR Labs (Meta), working on scaling multi-modal LLMs for vision, language and 3D representations. My representative works include **Pytorch3D** and **uCO3D** which have been open-sourced.

Prior to that, I was a research scientist at Amazon M5 Multimodal team on building VLMs for improving search relevance. My publications on **Amazon Science** reveal more details.

My research has been recognized with the JVCI 2021 Best Paper Award on interpretable CNNs, the IROS 2019 Best Paper Finalist on human–robot adversarial games (featured as a **USC headline**), and the CCBR 2016 Best Student Paper Award. I was fortunate to work with Prof. C.-C. Jay Kuo and Stefanos Nikolaidis at USC, Stan Z. Li, Shengcai Liao at the National Laboratory of Pattern Recognition (CASIA); Larry S. Davis at Amazon and David Novotny, Justin Johnson, Xinlei Chen at FAIR (Meta).

EDUCATION

2017 - 2021 **PhD in Electrical and Computer Engineering** — University of Southern California

2014 - 2017 **MSc in Computer Science** — Chinese Academy of Sciences

WORK EXPERIENCE (FULL LIST @LINKEDIN)

Senior Research Scientist — 12/2023 - Present
Apple's Foundational Diffusion Models — *Cupertino, CA*

- Lead end-to-end pre/post-training of **20B+ parameter audio–video generation foundation models** on **1K+ GPU** distributed training clusters; scope spans image, long-streaming video, joint autoregressive audio–video, and world models. Works under submission.
- Drive **post-training** research using **DPO, GRPO, and human-preference alignment**; design **agent post-training** pipelines where agents call the video/world models as tools to improve long-narrative coherence and audio–video synchronization, and to curate synthetic training data.
- Core contributor to **Image Playground and Genmoji**, shipped in Apple Intelligence.

Senior Research Scientist — 08/2023 - 12/2023
Graphics, Tencent America — *Palo Alto, CA*

Responsible for doing research on the latest development of generative AI that can help expedite the generation of 3D assets used in gaming. My role is to develop machine learning algorithms and models that facilitate 3d reconstruction given user's inputs, such as a 2D image or a textual description.

- Research on Image-to-3D Reconstruction for generic objects and characters in the game engine.
- Implement reconstruction pipeline based on 2D and 3D diffusion priors for 2D-lifting reconstruction models.

Research Scientist — 06/2022 - 07/2023
FAIR Labs, Meta AI — *Menlo Park, CA*

I work in FAIR Labs of Meta AI, responsible for fundamental research on foundation models in multi-modality and 3D. My role is mainly about developing large language models to push the boundary of AI.

- Member of Onevision dedicated to push the boundary of the next generation of foundational models, capable of visio-linguistic understanding in space/time and 2D/3D.

- Contributed to model-scaling of a 3D counterpart of Segment-Anything model, for learning category-agnostic 3D reconstruction priors. The work spans from SfM stack e.g., hloc, colmap integration, rendering and data curation to NeRF/Gaussian Splat, and text-to-3d generative models which culminated in **uCO3D** project.
- Contributor and maintainer for **Pytorch3d**. Contributed to the release of v0.7.1 and v0.7.2. Examples include Cuda/C++/python kernels for Marching Cubes, glTF utilities, and Fisheye Camera components.

Applied Scientist II 06/2021 - 06/2022
M5 Core Modeling, Search Science& AI, Amazon *Palo Alto, CA*

- Built **M5's LLM pretraining stack** for large-scale multi-modal pretraining of generic object representations, serving downstream services at Amazon (product search, CTR, multi-lingual/multi-task metrics); customized open-source frameworks (DeepSpeed, Hugging Face, Timm).
- *Why do we need large batch sizes in contrastive learning?* (NeurIPS).
- *Vision-Language Pre-Training with Triple Contrastive Learning* (CVPR).
- *SLADE: A Self-Training Framework for Distance Metric Learning* (CVPR).
- *Multi-Modal Alignment using Representation Codebook* (CVPR).

Research Assistant 09/2017 - 06/2021
University of Southern California *Los Angeles, CA*

PhD researcher working at the intersection of computer vision and robotics, supervised by Prof. C.-C. Jay Kuo and Prof. Stefanos Nikolaidis.

- *Interpretable Convolutional Neural Networks via Feedforward Design* (**JVCI 2021 Best Paper Award**).
- *Robot Learning via Human Adversarial Games* (**IROS 2019 Best Paper Finalist**); featured as a **USC headline**: "Showing Robots Tough Love Helps Them Succeed".
- *Human Decision Makings on Curriculum Reinforcement Learning with Difficulty Adjustment* (Arxiv 2022).
- *Metric-Aware Explainable Graph Network for Fashion Compatibility Recommendation* (Arxiv 2020).
- *PortraitGAN for Flexible Portrait Manipulation* (APSIPA 2020).

SELECTED PUBLICATIONS

JMLR 2025. Dongsheng Ding, Mihailo R. Jovanović, Kaiqing Zhang, **Jiali Duan**, Tamer Başar."Convergence and Sample Complexity of Natural Policy Gradient Primal-Dual Methods for Constrained MDPs".

CVPR 2025. Xingchen Liu, Piyush Tayal, Jianyuan Wang, Jesus Zarzar, Tom Monnier, Konstantinos Tertikas, **Jiali Duan**, Antoine Toisoul, Jason Y. Zhang, Natalia Neverova, Andrea Vedaldi, Roman Shapovalov, David Novotny."UnCommon Objects in 3D".

Apple ML Summit 2024. **Jiali Duan**, Vignesh Jagadeesh."Mitigating Person Hallucination in Diffusion Models under Constraints".

Arxiv 2022. Yilei Zeng, **Jiali Duan**, Yang Li, Emilio Ferrara, Lerrel Pinto, C.-C. Jay Kuo, Stefanos Nikolaidis. "Human Decision Makings on Curriculum Reinforcement Learning with Difficulty Adjustment".

CVPR 2022. **Jiali Duan***, Liqun Chen*, Son Tran, Jinyu Yang, Yi Xu, Zeng Belinda, Trishul Chilimbi. "Multi-modal Alignment using Representation Codebook"

CVPR 2022. Jinyu Yang, **Jiali Duan**, Son Tran, Liqun Chen, Yi Xu, Zeng Belinda, Trishul Chilimbi. "Multi-modal Representation Learning with Triple Contrastive Learning"

ICPR 2022. Xiaoyuan Guo*, **Jiali Duan***, C.-C. Jay Kuo, Judy Gichoya, Imon Banerjee. "Augmenting Vision Language Pretraining by Learning Codebook with Visual Semantics"

ICMR 2022. Xiaoyuan Guo, **Jiali Duan**, Saptarshi Purkayastha, Hari Trivedi, Judy Gichoya, Imon Banejee. "OSCARS: An Outlier-Sensitive Content-Based Radiography Retrieval System".

CVPR 2021. **Jiali Duan**, Yen-Liang Lin, Son Tran, Larry S. Davis, C.-C. Jay Kuo. "SLADE: A Self-Training Framework for Distance Metric Learning".

IROS 2019. **Jiali Duan***, Qian Wang*, Lerrel Pinto, C.-C. Jay Kuo, Stefanos Nikolaidis. "Robot Learning via Human Adversarial Games".

JVCI 2018. C.-C. Jay Kuo, Min Zhang, Siyang Li, **Jiali Duan**, Yueru Chen. Interpretable Convolutional Neural Networks via Feedforward Design.

ACM-TOMM 2017. **Jiali Duan**, Shuai Zhou, Jun Wan, Xiaoyuan Guo, Stan Z.Li. A Unified Framework for Multi-Modal Isolated Gesture Recognition.

ACCVW 2016. **Jiali Duan**, Jiali Duan, Shengcai Liao, Xiaoyuan Guo, Stan Z. Li. Face Detection by Aggregating Visible Components.

CCBR 2016. **Jiali Duan**, Shengcai Liao, Shuai Zhou, Stan Z. Li. Face Classification, A Specialized Benchmark Study.

AWARDS

- USC headline for IROS work on human–robot adversarial games, 2019.
- Presidential Outstanding Award, Chinese Academy of Sciences President Scholarship, 2017.
- 2nd Place, English Speaking Competition, University of Chinese Academy of Sciences, 2014.
- Honorable Mention, MCM/ICM Mathematical Contest in Modeling, 2013.
- Shanghai Regional Finalist, 21st Century Coca-Cola Cup National English Speaking Competition, 2012.
- Shanghai First Prize, Mathematical Modeling Contest, 2012.

PROFESSIONAL SERVICE/TALKS

- Associate Journal Editor for APSIPA
- Reviewer for CVPR, ECCV, ICCV, ICML, NeurIPS, ACL, EMNLP
- Multi-modal Alignment using Representation Codebook, CVPR 2022
- Technical talk at Intel 2022
- SLADE: A Self-Training Metric Learning Framework, CVPR 2021
- Guest Lecture for EE599-2020 (USC Deep Learning Course)
- Organizer/Presenter for USC Robotics Open House 2019

FEATURED WRITINGS

- Build a Simple LLM Agent from Scratch
- From DDPM to EDM: A Walkthrough of Diffusion Models
- Robotic Learning in Simulation
- Vision–Language Pretraining
- Face Detection by Aggregating Visible Components
- Face Classification: A Specialized Benchmark Study