

JIALI DUAN

jli.duan@gmail.com ◊ Linkedin ◊ Website

SUMMARY

I am a researcher with a strong track record in Computer Vision and Artificial Intelligence. I am particularly interested in **multi-modality learning**, **foundation models**, and **generative AI**. I am into the research for how these models fundamentally work but also hands on production stacks within the leading industry standards to marriage technology with products. My latest research and engineering focus is on multi-modal generative modelling involving generation and understanding of languages, images, and videos.

EDUCATION

2017 - 2021 **PhD in Electrical and Computer Engineering** University of Southern California
2014 - 2017 **MSc in Computer Science** Chinese Academy of Sciences
2010 - 2014 **BS in Information Engineering** East China University of Science and Technology

WORK EXPERIENCE

Multimodal Generative Modeling Researcher 12/2023 - Present
Machine Learning, Apple *1 Apple Park Way, Cupertino, CA*

Research and production in the domain of diffusion and multimodal large-language models. I'm interested in

- Applying multimodal foundational models to facilitate user experience.
- Content generation with AI and efficient on-device deployment.

Senior Researcher 08/2023 - 12/2023
Graphics, Tencent America *2747 Park Blvd, 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 *1 Hacker Way, Menlo Park, CA*

I work in FAIR Labs of Meta AI, responsible for doing research on AI and publishing academic research papers. 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 data-scaling and model-scaling efforts of a 3D counterpart of Segment-Anything model, for learning category-agnostic 3D reconstruction priors.
- Contributor and Reviewer 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.
- Lead the 3D data crowd-sourcing efforts by deploying Amazon Mturk websites and building the structure-from-motion stack including hloc, colmap integration, 3D animation and selection criteria metrics.

Applied Scientist II 06/2021 - 06/2022
M5 Core Modeling, Search Science& AI, Amazon *611 Cowper St, Palo Alto, CA*

Applying AI technologies to make recommendation algorithms more intelligent in understanding users' needs. I develop recommendation machine learning models to deliver product recommendations to the users.

- Performed large multi-modal pretraining for generic object embeddings that serve a variety of downstream services at Amazon including product search, click-through-rate (CTR) and improving multi-lingual, multi-task metrics.
- Contributed to the LLM infra stack by customizing and diving into open-source frameworks such as DeepSpeed, Hugging Face and Timm.
- Mentored two interns and published 1 NeurIPS, 2 CVPR papers on multi-modal representation learning.

Research Assistant

University of Southern California

08/2017 - 05/2021

Los Angeles, CA

- Research on interpretable deep neural networks with feed-forward design. Awarded JVCi 2021 Best Paper.
- Proposed a pioneering framework that insinuates the idea of adversarial learning with human robot interaction that hits USC headline and IROS 2019 Best Paper Finalist.
- Research and engineering on interactive high-resolution portrait manipulation using generative adversarial networks (GANs).

SELECTED PUBLICATIONS

NeurIPS 2022. Changyou chen, Jianyi Zhang, Yi Xu, Liqun Chen, **Jiali Duan**, Yiran Chen, Son Tran, Belinda Zeng, Trishul Chilimbi. “Why do We Need Large Batchesizes in Contrastive Learning? A Gradient-Bias Perspective”.

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. “OS-CARS: An Outlier-Sensitive Content-Based Radiography Retrieval System”.

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

SCMLS 2020. **Jiali Duan**, Xiaoyuan Guo, Son Tran, C.-C. Jay Kuo. “Fashion Compatibility Recommendation via Unsupervised Metric Graph 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.

PROFESSIONAL SERVICE

- Associate Journal Editor for APSIPA
- Reviewer for CVPR, ECCV, ICCV, ICML, NeurIPS, ACL, EMNLP, TOMM, RA-L, ICIP
- Presenter for USC Robotics Open House 2019

AWARDS

- Best Paper Award in the 2021 Journal of Visual Communication and Image Processing.
- Best Paper Award in the 2019 International Conference on Intelligent Robots and Systems.
- Best Student Paper Award for CCBR 2016.