Fanzi Wu

Senior Applied Scientist — Multimodal Video & Image Generation | Amazon AGI

I focused on multimodal image and video generation, cross-modal alignment, with 10+ years experience advancing model capability, data quality, and production systems.

email: fzwu@link.cuhk.edu.hk **phone**:(+1) 206-886-4479

Education

The Chinese University of Hong Kong **8/2014–3/2020**

Ph. D. in Electronic Engineering, advised by Prof. King Ngi Ngan and Prof. Thierry Blu

Tianjin University **9/2010–7/2014**

B. Eng. in Electronic Engineering

Experience

Senior Applied Scientist @ Amazon AGI **8/2024–Present**

- **Talking Head video generation:** Led the end-to-end development of a cross-modal talking-head video generation system: including building a **large-scale data curation pipeline** based on **PySpark** and **Ray**, delivering 10M+ training set with aligned audio, text, and videos; fine-tuning model to improve lip-sync accuracy, identity consistency, and temporal stability; and deploying a **real-time model** to support a streaming interactive visual chatbot.
- **Omni video generation:** Led **post-training and alignment** for omni video generation models, exploring late-fusion architectures that combine LLMs with diffusion-based video models to enhance multimodal reasoning and prompt adherence, and driving model-level improvements in conditioning, guidance strategies, and instruction following to support robust, production-scale video generation.
- **Omni image generation:** Core contributor to **large-scale autoregressive image generation model**, developing and optimizing training recipes including data mixtures and curriculum strategies, and delivering modeling and inference-time scaling enhancements that improved generation fidelity and quality–compute tradeoffs.

Applied Scientist @ Amazon Web Service **8/2021–8/2024**

- Face Recognition: Owned and shipped **head pose** and **gaze estimation models** for the **Amazon Rekognition Face API** (link), conducting research on **data efficiency and robustness under long-tailed distributions** at **100M+ data scale**.
- Bedrock Data Automation: **Led the open-set logo detection project**, overcoming data annotation challenges through **ML-assisted pseudo-labeling** and launching the model as a Bedrock Data Automation capability (link).

Applied Researcher @ Tencent IEG **9/2020–7/2021**

- **Customized Avatar Creation:** Delivered an **image-based avatar creation system** for *QQ Speed Drifter* (link), introduced **few-shot learning techniques** to significantly reduce reliance on manual character artist workflows while maintaining visual quality and stylistic consistency.

Intern @ Amazon Rekognition **11/2019-1/2020**

- Developed face landmark tracking system for video with temporal stability.

Intern @ Tencent AI Lab **6/2018-6/2019**

- Research work on face reconstruction from multiple view images, the work is published in a CVPR 2019 paper (link).

Selected Publication

-Autoregressive Refinement of Text-to-Image Generation through Iterative Scale Enhancement

This work introduced training and inference scaling, and scale experts to improve autoregressive image generation.

Work In Progress, 2025

-PartMesh: Image-to-Mesh with Part Information via Auto-regressive Transformer

PartMesh is an autoregressive, part-based 3D mesh generation model that improves topology and semantic structure over existing methods.

Submitted to US Patent, 2024

-Learning Facial Landmarks Detection from Multi-domain Synthetic Faces

This work proposed a CLIP-guided, synthetic-only landmark detection framework with interleaved adapters and dual heads for robust 2D/3D face alignment across domains.

Submitted to US Patent, 2023

-Privacy Attacks and Privacy Preserving Training of Deep Metric Learning

This work studies privacy risks in deep metric learning via identity inference attacks and proposes differential privacy–based defenses.

Submitted to US Patent, 2022

-Self-Supervised Learning of Detailed 3D Face Reconstruction

Yajing Chen, **Fanzi Wu**, Zeyu Wang, Yibing Song, Yonggen Ling, Linchao Bao

IEEE Transactions on Image Processing, 2020

-Mvf-net: Multi-view 3d face morphable model regression

Fanzi Wu, Linchao Bao, Yajing Chen, Yonggen Ling, Yibing Song, Songnan Li, King Ngi Ngan, Wei Liu

IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2019

-Cascaded regression using landmark displacement for 3D face reconstruction

Fanzi Wu, Songnan Li, Tianhao Zhao, King Ngi Ngan, Lu Sheng

Pattern Recognition Letters, 2019

-3-D reconstruction of human body shape from a single commodity depth camera

Tianhao Zhao, Songnan Li, King Ngi Ngan, **Fanzi Wu**

IEEE Transactions on Multimedia, 2018

-A facial expression model with generative albedo texture

Songnan Li, **Fanzi Wu**, Tianhao Zhao, Ran Shi, King Ngi Ngan

Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA), 2016

-Model-based face reconstruction using SIFT flow registration and spherical harmonics

Fanzi Wu, Songnan Li, Tianhao Zhao, King Ngi Ngan

International Conference on Pattern Recognition (ICPR), 2016

Skills

Programming: Python, C++, MATLAB

ML Frameworks: PyTorch, PySpark, Ray