Kangfu Mei, Ph.D

CONTACT

1600 Amphitheatre Pkwy
Mountain View, CA 94043
United States

Tel: (+1) 443-240-5261
Email: mikumkf@gmail.com

CURRENT

- Google DeepMind – Senior Research Scientist — Mountain View, CA
  Work responsibilities: Technical lead for Video RL research, developing reinforcement learning and adaptive reward modeling methods for human preference alignment in video generation. Core contributor to Veo 3, Genie 3, and Gemini Omni Flash.

EXPERIENCE

- Google Research / Google — Mountain View, CA
  Student Researcher / Research Scientist — 05/2023 - 06/2025
- Adobe Research, Research Engineering and Design Lab (RED) — San Jose, CA
  Research Intern — 05/2022 - 11/2022
- Alibaba-Group, DAMO Academy — Shenzhen, China
  Research Intern — 06/2020 - 11/2020
- Kuaishou (Kling) Technology — Beijing, China
  Imaging Algorithm Engineer Intern — 07/2018 - 05/2019

EDUCATION

- Johns Hopkins University — Baltimore, MD
  Ph.D. Department of Electrical and Computer Engineering — 09/2021 - 12/2024
  Doctoral Committee: Prof. Vishal M. Patel & Rama Chellappa & Alan Yuille
- The Chinese University of Hong Kong — Shenzhen, China
  M.Phil. School of Science and Engineering — 09/2019 - 06/2021
- Jiangxi Normal University — Nanchang, China
  B.Eng. School of Computer Science and Engineering — 09/2015 - 06/2019

PROJECTS

- Gemini Omni Flash deepmind.google/models/gemini-omni/ — 2025 – 2026
  As technical lead for Video RL, I set the research direction and coordinate efforts within the team and with reward modeling and other partner teams to improve human preference alignment and factual accuracy in video generation. To address policy lag from long video rollouts in asynchronous training, I proposed a single-step, DPO-inspired preference optimization method that avoids backpropagation through full denoising trajectories, with a clipped trust-region constraint to stabilize updates. Together with the disaggregated RL infrastructure I designed and built, this enabled scalable full-model training for long-form video generation. With the reward modeling team, I also formulated an adaptive preference modeling approach to track evolving user preferences. These contributions helped Gemini Omni Flash maintain the highest Elo rating on the Arena AI video-generation leaderboard from its release to date.
- Veo 3.1 & Genie 3 deepmind.google/models/veo/ & deepmind.google/models/genie/ — 2025
  As a core contributor to Veo 3.1 and Genie 3, I investigated how generative models can better follow user intent and respond to actions. For Veo 3.1, I designed controlled architecture studies and developed diffusion guidance and preference-based post-training methods to improve generation quality and controllability. For Genie 3, I developed reinforcement learning methods to improve action-control accuracy, addressing whether generated world dynamics respond correctly to the actions supplied to the model.

PUBLICATIONS Google Scholar Profile: https://scholar.google.com/citations?user=e_nu_TIAAAAJ&hl=en (Oct 2026) Citations: 2577 H-Index: 17

SELECTED RESEARCH:

[PDF] [arXiv] [Github] Kangfu Mei, Vishal M Patel. *"VIDM: Video Implicit Diffusion Models"* AAAI Conference on Artificial Intelligence (AAAI), Oral, 2023.
Among the earliest diffusion-based video generation methods (2022), outperforming then-leading GAN-based approaches.

[arXiv] Kangfu Mei, Zhengzhong Tu, Mauricio Delbracio, Hossein Talebi, Vishal M. Patel, Peyman Milanfar. *"Bigger is not Always Better: Scaling Properties of Latent Diffusion Models"* Transactions on Machine Learning Research, 2025.
Extends scaling to account for diffusion guidance and sampling budgets, with the first 5B-parameter latent diffusion model trained in 2023.

[arXiv] [Code] Kangfu Mei, Mo Zhou, Vishal M. Patel. *"Field-DiT: Diffusion Transformer on Unified Video, 3D, and Game Field Generation"* International Conference on Learning Representations (ICLR), 2025.
Recasts video, 3D scenes, and interactive games as generative fields within a unified diffusion modeling framework.

CONFERENCE PAPERS: (3 ICLR& NeurIPS & ICML, 6 CVPR & ECCV, 2 AAAI, 2 WACV, 1 ACCV)

[arXiv] [C01] Yuyang Hu, Mojtaba Sahraee-Ardakan, Kangfu Mei, Arpit Bansal, Christian Qi, Peyman Milanfar, Mauricio Delbracio *"Generative Manifold Distillation: Aligning Restoration Trajectories with the Natural Image Prior "* European Conference on Computer Vision (ECCV), 2026.

[arXiv] [C02] Jinxiu Liu, Jianru Li, Tanqing Kuang, Xuanming Liu, Kangfu Mei, Yandong Wen, Weiyang Liu *"SymbOmni: Evolving Agentic Omni Models via Symbolic Concept Learning "* European Conference on Computer Vision (ECCV), 2026.

[arXiv] [C03] Jinxiu Liu, Xuanming Liu, Kangfu Mei, Yandong Wen, Weiyang Liu *"XYZFlow: Scaling Multidimensional Shortcut Flows for Efficient Generative Modeling "* International Conference on Machine Learning (ICML), 2026

[arXiv] [C04] Yuyang Hu, Kangfu Mei, Mojtaba Sahraee-Ardakan, Ulugbek S Kamilov, Peyman Milanfar, Mauricio Delbracio *"Kernel Density Steering: Inference-Time Scaling via Mode Seeking for Image Restoration "* Annual Conference on Neural Information Processing Systems (NeurIPS), 2026

[arXiv] [C05] Kangfu Mei, Hossein Talebi, Mojtaba Ardakani, Vishal M. Patel, Peyman Milanfar, Mauricio Delbracio. *"The Power of Context: How Multimodality Improves Image Super-Resolution"* IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2025.

[arXiv] [C06] Kangfu Mei, Mo Zhou, Vishal M. Patel. *"Field-DiT: Diffusion Transformer on Unified Video, 3D, and Game Field Generation"* International Conference on Learning Representations (ICLR), 2025.

[PDF] [arXiv] [Github] [C07] Kangfu Mei, Mauricio Delbracio, Hossein Talebi, Zhengzhong Tu, Vishal M

Patel, Peyman Milanfar. *"CoDi: Conditional Diffusion Distillation for Higher-Fidelity and Faster Image Generation"* IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2024.

[PDF] [arXiv] [Github] [C08] Kangfu Mei, Luis Figueroa, Zhe Lin, Zhihong Ding, Scott Cohen, Vishal M. Patel. *"Latent Feature-Guided Diffusion Models for Shadow Removal"* IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), 2024.

[PDF] [arXiv] [Github] [C09] Kangfu Mei, Vishal M Patel. *"VIDM: Video Implicit Diffusion Models"* AAAI Conference on Artificial Intelligence (AAAI), Oral, 2023.

[PDF] [arXiv] [Github] [C010] Nithin Gopalakrishnan Nair, Kangfu Mei, Vishal M Patel. *"AT-DDPM: Restoring Faces degraded by Atmospheric Turbulence using Denoising Diffusion Probabilistic Models"* IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), 2023.

[PDF] [arXiv] [Github] [C011] Kangfu Mei, Vishal M Patel, Rui Huang. *"Deep Semantic Statistics Matching (D2SM) Denoising Network "* European Conference on Computer Vision (ECCV), 2022.

[PDF] [arXiv] [Github] [C012] Kangfu Mei, Shenglong Ye, Rui Huang. *"SDAN: Squared Deformable Alignment Network for Learning Misaligned Optical Zoom"* IEEE International Conference on Multimedia and Expo (ICME), 2021.

[PDF] [arXiv] [Github] [C013] Qi Song, Kangfu Mei, Rui Huang. *"AttaNet: Attention-augmented network for fast and accurate scene parsing"* AAAI conference on artificial intelligence (AAAI), 2021.

[PDF] [Github] [C014] Juncheng Li, Yiting Yuan, Kangfu Mei, Faming Fang. *"Lightweight and Accurate Recursive Fractal Network for Image Super-Resolution"* IEEE/CVF International Conference on Computer Vision Workshop (ICCVW), 2019.

[PDF] [C015] Kangfu Mei, Juncheng Li, Jiajie Zhang, Haoyu Wu, Jie Li, Rui Huang. *"Higher-resolution network for image demosaicing and enhancing"* IEEE/CVF International Conference on Computer Vision Workshop (ICCVW), 2019.

[PDF] [Github] [C016] Juncheng Li, Faming Fang, Kangfu Mei, Guixu Zhang. *"Multi-scale Residual Network for Image Super-Resolution"* European Conference on Computer Vision (ECCV), 2018.

[PDF] [Github] [C017] Kangfu Mei, Aiwen Jiang, Juncheng Li, Mingwen Wang. *"Progressive feature fusion network for realistic image dehazing"* Asian Conference on Computer Vision (ACCV), 2018.

JOURNAL ARTICLES: (1 JSTSP, 1 TCSVT, 2 TMLR)

[arXiv] [J01] Mo Zhou, Keren Ye, Viraj Shah, Kangfu Mei, Mauricio Delbracio, Peyman Milanfar, Vishal M. Patel, Hossein Talebi
"Reference-Guided Identity Preserving Face Restoration" Transactions on Machine Learning Research (TMLR), 2026

[arXiv] [J02] Kangfu Mei, Zhengzhong Tu, Mauricio Delbracio, Hossein Talebi, Vishal M. Patel, Peyman Milanfar. *"Bigger is not Always Better: Scaling Properties of Latent Diffusion Models"* Transactions on Machine Learning Research (TMLR),

2025.

[PDF] [arXiv] [J03] Kangfu Mei, Vishal M. Patel. *"Ltt-gan: Looking through turbulence by inverting gans"* IEEE Journal of Selected Topics in Signal Processing (JSTSP), 2023.

[PDF] [arXiv] [J04] Juncheng Li, Faming Fang, Jiaqian Li, Kangfu Mei, Guixu Zhang. *"MDCN: Multi-scale Dense Cross Network for Image Super-Resolution"* IEEE Transactions on Circuits and Systems for Video Technology (TCSVT), 2020.

ACTIVITIES

- Reviewer / Area Chair for International Conferences
  - AC of Winter Conf. on Applications of Computer Vision (WACV) 2025 – 2026
  - IEEE Conf. on Computer Vision and Pattern Recognition (CVPR) 2020 – 2026
  - International Conf. on Computer Vision (ICCV) 2021 – 2026
  - European Conf. on Computer Vision (ECCV) 2020 – 2026
  - AAAI Conf. on Artificial Intelligence (AAAI) 2021 – 2025
  - Winter Conf. on Applications of Computer Vision (WACV) 2021 – 2024
  - Asian Conf. on Computer Vision (ACCV) 2018 – 2024
- Reviewer for International Journals
  - IEEE Trans. on Neural Networks and Learning Systems (TNNLS) 2022
  - IEEE Trans. on Circuits and Systems for Video Technology (TCSVT) 2022
  - IEEE Trans. on Image Processing (TIP) 2022
  - IEEE Trans. on Multimedia (TMM) 2023
  - International Journal of Computer Vision (IJCV) 2023 – 2024
  - Computer Vision and Image Understanding (CVEU) 2021 – 2022

PRESENTATIONS

The Power of Inference Time Scaling for Multimodal Generative Foundation Models, invited talk at National Taiwan University (NTU). (Jul 2026)

The Power of Context: How Multimodality Improves Computational Photography, invited talk at ICLR 2025, Google Booth. (May 2025)

The Power of Context: How Multimodality Improves Computational Photography, invited talk at NUS Open Multimodal Gathering workshop. (May 2025)

Deep Generative Models and Computational Photography, Luma Seminar, Google. (Jun 2023)

Conditional Diffusion Distillation for Higher-Fidelity and Faster Image Generation, CCI CVPR Share-a-thon, Google. (Dec 2023)

Video Implicit Diffusion Models, AAAI23 Pre-presentation, AI TIME. (Jan 2023)

HONORS

- First place, Advances in Image Manipulation Challenges (RAW2RGB) at ICCV 2019
- 6th, New Trends in Image Restoration and Enhancement (Dehazing) at CVPR 2018