David Fan

davidfan219@gmail.com | https://davidfan.io | github.com/dfan | linkedin.com/in/davidfan97

I do foundational research in multimodal representation learning and video understanding that pushes state-of-the-art and generates product impact.

Select Publications

Select Peer-Reviewed

1. **David Fan**†, Shengbang Tong†, Jiachen Zhu, Koustuv Sinha, Zhuang Liu, Xinlei Chen, Michael Rabbat, Nicolas Ballas, Yann LeCun, Amir Bar, Saining Xie. Scaling Language-Free Visual Representation Learning. *ICCV*, 2025 (*Highlight*). **Self-supervised learning can match CLIP-pretraining at scale with enough diverse data.**
2. Shengbang Tong, **David Fan**, Jiachen Zhu, Yunyang Xiong, Xinlei Chen, Koustuv Sinha, Michael Rabbat, Yann LeCun, Saining Xie, Zhuang Liu. MetaMorph: Multimodal Understanding and Generation via Instruction Tuning. *ICCV*, 2025. **Visual understanding and generation are mutually beneficial in unified multimodal models.**
3. **David Fan**, Jue Wang, Shuai Liao, Yi Zhu, Vimal Bhat, Hector Santos Villalobos, Xinyu Li. Motion-Guided Masking for Spatiotemporal Representation Learning. *ICCV*, 2023.
Self-supervised masking alg. for video masked autoencoder achieves SOTA with 3x less pretraining.

Tech Reports

1. JEPA Team. V-JEPA 2: Self-Supervised Video Models Enable Understanding, Prediction and Planning. **Self-supervised learning from video at scale leads to physical understanding, prediction, and planning.**

Professional Experience

Meta Fundamental AI Research (FAIR) — *New York, NY*

Senior Research Engineer — *2024 - current*

- Co-led comprehensive empirical study into the science behind unified multimodal pretraining. Produced proof-of-concept for unifying world modeling capabilities with multimodal models via actions as language and visual latent space prediction. Paper submitted to ICML.
- Core contributor to V-JEPA 2, a self-supervised video model trained on >1M hours of video, with physical understanding, prediction, and planning capabilities. Built 8B param multimodal language model upon V-JEPA 2, that achieves SOTA (in model class) on video question answering tasks, e.g. PerceptionTest. >1M downloads.
- [ICCV 2025 *Highlight* (1st auth)] WebSSL is the first self-supervised vision encoder that achieves performance parity with CLIP on multimodal tasks. We demonstrate that self-supervised learning scales better than CLIP with appropriate diversity and scale of data. >100K HuggingFace downloads.
- [ICCV 2025 (2nd auth)] MetaMorph is a unified autoregressive model capable of understanding and generating both text and visual tokens. We demonstrate that visual understanding and generation capabilities are mutually beneficial, and that unified multimodal models can leverage world knowledge + reasoning during visual generation.

Amazon Prime Video — *Seattle, WA and New York, NY*

Applied Research Scientist — *2020 - current*

- [Prod] Trained 1B param multimodal foundation model with large-scale vision-language-audio pretraining. Outperforms OpenAI CLIP by 25% on internal zero-shot classification and retrieval benchmarks.
- [Prod] Enabled automated video advertisement insertion (CEO-level goal) with novel video segmentation model.
- [Prod] Developed embeddings for visual search and recommendation which outperform baseline recsys by 5%.
- [Prod] Designed transformers with multimodal fusion for automated content moderation and compliance.
- [Prod] Built distributed PyTorch training codebase and managed compute infrastructure for larger org of 30+ ICs.
- [ICCV 2023 (1st auth)] New SOTA video masked autoencoders by 5% for action recognition. Amazon blog.
- [ICLR 2023 (1st auth)] Nearest-neighbor sampling improves positive pair diversity for video contrastive learning.
- [CVPR 2021] New SOTA movie segmentation by 13% while reducing annotation costs by 75% ($200K/yr). Blog.
- Mentored two research interns to full-time offer.

Amazon Web Services — *Seattle, WA*

Research Engineer — *2019 - 2020*

- Launched GPU slicing service for PyTorch inference on AWS SageMaker and EC2 (blog post).
- Reduced inference latency by up to 70% via mixed precision, TensorRT optimization, and optimizing data loading.

Princeton Vision and Learning Lab (Prof. Jia Deng) *Princeton, NJ*

Researcher

- Created OASIS - the first dataset for single-image 3D vision in the wild with dense annotations of detailed 3D geometry at scale. Dataset improves performance in multiple single-image 3D tasks.
- Implemented novel annotation pipeline for crowdsourcing dense pixel-wise 3D ground truths from sparse annotations. Implemented annotation quality control and reward workflows.
- Trained state-of-the-art models for monocular surface normal estimation (hourglass network), planar semantic segmentation (DeepLab), fold and occlusion boundary detection (HED). Evaluated downstream generalization.
- [CVPR 2020 Paper]. Senior thesis won CS department Sigma Xi award.

Education

Princeton University *2015 - 2019*

B.S.E in Computer Science *Magna Cum Laude (High Honors)*

Minor in Statistics and Machine Learning

Thesis Advisor: Prof. Jia Deng

Invited Talks

Scaling Language-Free Representation Learning

- ICCV Workshop on Knowledge-Intensive Multimodal Reasoning (October 2025)
- Cohere Labs (August 2025)
- Oxford VGG Seminar (May 2025)

Motion-Guided Masking for Spatiotemporal Learning

- Amazon Foundation Model Symposium (December 2023)

From Bioinformatics to Machine Learning

- National Science Olympiad Tournament (May 2022)

Building Computer Vision Models with Limited Labeled Data

- Amazon Research (March 2021)

3D Surfaces in the Wild

- Princeton Research Day (May 2019)
- Princeton Computer Science Independent Work Poster Session (May 2019)

Visualizing geographic trends in insurance claims data

- Harvard Medical School DBMI Summer Symposium (July 2016)

Honors & Awards

Nominated to Princeton University Alumni Council Executive Committee	2024
President of Princeton Club of Western Washington	2021-2023
Kaggle Bronze Medal (Google Open Images – Object Detection)	2019
Sigma Xi Award for Outstanding Undergraduate Research, Princeton University	2019
Class of 1901 Medal Finalist, Princeton University	2019
Graduating senior who has done the most for Princeton University. Princeton Innovation Magazine 25 under 25	2016
Intel Science Talent Search Semifinalist	2015

Service

Peer-Reviewing

- CVPR (2023, 2024, 2025), ICCV (2023, 2025), ECCV (2024, 2026), ICML (2026)

STEM Education

- Asian-American Scholar Forum (2024-2025)
- Washington State Science Olympiad Supervisor (2020-2021)
- NJ State Science Olympiad Supervisor (2018-2019)
- MIT Science Olympiad Supervisor (2016)

Leadership

Princeton Club of Western Washington
Seattle, WA

President *Oct. 2021 - Oct. 2023*

- President of 2,000-person regional Princeton alumni association (8th largest). Organized social and networking events and helped coordinate alumni interviewing for prospective students.

HackPrinceton
Princeton, NJ

Director *Sept. 2016 - Apr. 2018*

- Led 30 organizers and raised $130,000 in funding as head director of HackPrinceton Fall 2017 and Spring 2018, which hosted 1,100 students from around the world.
- Organized logistics and hacker experience for Fall 2016 and Spring 2017.
- Developed competition website: https://f17.hackprinceton.com

Princeton University Science Olympiad Tournament
Princeton, NJ

Co-founder *Sept. 2016 - Feb. 2019*

- Founded one of the USA's premier high school science competitions. Hosts 800 high school students annually.
- Directed a team of 20 organizers and 100+ volunteers to organize 23 competition events.
- Created website and built organization brand from ground-up: https://scioly.princeton.edu

Princeton University Math Competition
Princeton, NJ

Logistics Director *Sept. 2016 - Nov. 2016*

- Directed logistics for one of the nation's premier high school math competitions.