David Fan

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

Education

Princeton University *Sept. 2015 - June 2019*

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

Minor in Statistics and Machine Learning

- *Thesis Advisor:* Prof. Jia Deng
- *Select Courses:* Graduate Computer Vision, Machine Learning, Optimization, Probability and Stochastic Systems, Theory of Algorithms, Algorithmic Game Theory, Information Security, Functional Programming

Publications

Peer-Reviewed

1. **David Fan**, Jue Wang, Shuai Liao, Yi Zhu, Vimal Bhat, Hector Santos-Villalobos, Rohith MV, Xinyu Li. Motion-Guided Masking for Spatiotemporal Representation Learning. *International Conference on Computer Vision (ICCV)*. 2023.
2. **David Fan**, Deyu Yang, Xinyu Li, Vimal Bhat. Nearest-Neighbor Contrastive Learning from Unlabeled Videos. *International Conference on Learning Representations (ICLR) Workshop on Mathematical and Empirical Understanding of Foundation Models*. 2023.
3. Shixing Chen, Xiaohan Nie†, **David Fan**†, Dongqing Zhang, Vimal Bhat, Raffay Hamid. Shot Contrastive Self-Supervised Learning for Scene Boundary Detection. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2021.
4. Weifeng Chen, Shengyi Qian, **David Fan**, Noriyuki Kojima, Max Hamilton, Jia Deng. OASIS: A Large-Scale Dataset for Single Image 3D in the Wild. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2020.
5. Zachary A Kopp, Jo-Lin Hsieh, ..., **David Fan**, ..., Yongkyu Park. Heart-specific Rpd3 Downregulation Enhances Cardiac Function and Longevity. *Aging*. 2015.
6. Shi-Zeng Lin, Xueyun Wang, ..., **David Fan**, ..., Sang-Wook Cheong. Topological defects as relics of emergent continuous symmetry and Higgs condensation of disorder in ferroelectrics. *Nature Physics*. 2014.

†denotes equal contribution.

Research / Work Experience

Amazon *Seattle, WA and New York, NY*

Applied Scientist *July 2020 - current*

- Developed self-supervised method for scene boundary detection that improves state-of-the-art by 13% while reducing data labeling requirements by 75% and improving inference time by 84%. Second-author paper accepted to CVPR 2021. Blog post on Amazon Research website.
- Delivered automatic advertisement insertion feature built on top of my scene boundary model. This work had direct CEO visibility and enabled Prime Video's north star to become a marketplace for AVOD content. See launch announcement.
- Developed "Motion-Guided Masking" which improves video representation learning by conditioning masked video pretraining with motion-guided masks. Leveraged motion-vectors to achieve efficient motion guidance directly from MP4 video. Up to 5% better performance on video action recognition benchmarks. First-author paper paper accepted to ICCV 2023.
- Developed multimodal Transformers with audio, text, and video for content moderation system that automatically detects sensitive explicit scenes in movies.
- Built multi-GPU distributed PyTorch codebase for training large multimodal foundation models.
- Mentored two research interns to full-time offer.

Amazon Web Services — *Seattle, WA*
Software Engineer — *Aug. 2019 - July 2020*

- Added logging metrics and launched canaries to support new EC2 G4 instance family based accelerators.
- Launched Elastic Inference-enabled PyTorch (blog post) for SageMaker, EC2, ECS.
- Implemented TorchScript graph validation, shipped updated AWS Deep Learning Conda environments and Docker containers, benchmarked performance, wrote blog post.
- Created proof-of-concept for building and integrating TensorRT-enabled TensorFlow 2.1 into the inference engine. Reduced latency by up to 70% compared to FP32 native TensorFlow in benchmarks.

Princeton Vision and Learning Lab (Prof. Jia Deng) — *Princeton, NJ*
Undergraduate Researcher — *Sept. 2018 - July 2019*

- Created OASIS, which is 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 pipeline for crowdsourcing dense pixel-wise 3D ground truths from sparse annotations, and quality control mechanisms.
- Trained state-of-art deep learning models to benchmark dataset for monocular surface normal estimation and planar semantic segmentation, and evaluate cross-dataset generalization. Provided baseline for fold and occlusion boundary detection.
- Paper accepted to CVPR 2020.

Amazon Web Services — *East Palo Alto, CA*
Software Engineering Intern — *June 2018 - Aug. 2018*

- Developed automated devops tool for AWS Aurora — a distributed cloud-native relational database service — which improved on-call engineer productivity by automatically applying fixes to low-severity tickets and reducing manual processes.
- Wrote tool for applying autoscaling policies and provisioning IOPS for DynamoDB clusters to improve cost management.

Phosphorus — *New York, NY*
Software Engineering Intern — *May 2017 - Aug. 2017*

- Redesigned and implemented custom UI/UX components for user dashboard using Wicket and Scala.
- Designed modeling layer in Scala, Spring Boot, Hibernate, and PostgreSQL. Wrote AWS CloudFormation templates for automated infrastructure deployment.

Harvard-MIT HST (Biomedical Informatics) — *Boston, MA*
Research Intern — *June 2016 - Aug. 2016*

- Developed web app for visualizing geographic trends in AETNA insurance and US Census data using R Shiny and MySQL. Mentored by Prof. Isaac Kohane and Prof. Arjun Manrai.
- Contributed to ubit2.com, an open-source client-side web application for bioinformatic analyses. Technical report on bioRxiv.

Leadership

Princeton Club of Western Washington — *Seattle, WA*
President — *Oct. 2021 - present*

- President of regional Princeton alumni association with 2,000 active members. Organized social and networking events and helped coordinate alumni interviewing for prospective students.

HackPrinceton — *Princeton, NJ*
Director — *Sept. 2016 - Apr. 2018*

- Princeton's biannual hackathon hosts 1,100 students from around the world each year. Led 30 organizers and raised $130,000 in funding as head director of HackPrinceton Fall 2017 and Spring 2018.
- Organized logistics and hacker experience for Fall 2016 and Spring 2017.
- Past website: https://f17.hackprinceton.com

Princeton University Science Olympiad *Princeton, NJ*

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

- 800 of the USA's top high school students compete at the annual Princeton University Science Olympiad invitational tournament.
- Founded organization in 2016 and directed a team of 10 organizers + 100 volunteers to run the inaugural tournament. Coordinated 23 competition events and over 100 student volunteers.
- In February 2018, became first tournament nationwide to waive registration fees and release all tests, improving accessibility for underresourced groups.
- Created website and organizational presence: 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.

Select Software Projects

Single-Image Normal Estimation *2018*

- Implemented hourglass network architecture from NeurIPS 2016 paper and trained on internal dataset for Princeton competition. Used data augmentations and learning rate scheduling to win 1st place.
- Code: https://github.com/dfan/single-image-surface-normal-estimation

TigerTexts *2018*

- Web app that consolidates Princeton student coursebook pricing information from multiple sources and offers third-party seller platform. Uses the MERN (MongoDB, Express, React.js, Node.js) stack.
- Documentation and technical report: https://tigertexts.herokuapp.com/about
- Code: https://github.com/rfblue2/tigertexts

UBiT2 *2016*

- Lightweight client-side web app for visualization + analysis of RNA-seq + qPCR data. All computation is client-side enabling greater accessibility for researchers without programming background. No setup required.
- Link: ubit2.com | Code: https://github.com/JEFWorks/ubit2

Honors & Awards

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

Talks

Building CV models with limited labeled training 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)

Skills

Programming Languages	Python, Java, R, Javascript, Go, OCaml, MatLab, C/C++
Libraries	PyTorch, TorchScript, TensorFlow, OpenCV, Shiny
Web Development	Django, Express.js | HTML5, React, Hugo, Jekyll
Databases	MySQL, MongoDB, DynamoDB
Other	AWS, Docker, Git, UNIX, LaTeX, Leadership