ORR ZOHAR

orrzohar@stanford.edu ⋄ orrzohar.github.io

EXPERIENCE

PhD Researcher, Stanford University *Sep 2021 - Present*

MARVL, SVL; *Large Multimodal Models; Video Understanding.* Working on video Large Multimodal Models (video-LMMs), focusing on self-training, agentic design, and architecture modification for long-context reasoning and efficient test-time compute scaling.

- Developed Video-STaR, the first video-LMM self-training method that generates chain-of-thought video instruction tuning datasets by leveraging weak supervision (in collaboration with Google Research).
- Helped create VideoAgent, the first video-LMM agent method that utilizes reflection and retrieval to achieve strong and efficient long-video performance.
- Developed a vision-based surgical training platform enabling students to upload practice videos, which a custom multi-task model evaluates for surgical skill.
- Developed robust object detection models in the open-world setting, where models are expected to detect and incrementally learn unknown objects and introduced foundation models to open-world object detection.

Meta - GenAI *Jun 2024 - Dec 2024*

Research Scientist Intern at Multimodal Foundations Team

- Led Apollo, a large study exploring video-Large Multimodal Models (LMMs), trained hundreds of video-LMMs and showed what video-specific design choices affect video understanding performance.
- Spearheaded the Multimodal Yarn project, generalizing models from short to long videos through interleaved pre-training on multimodal datasets.
- Developed a transformer-based video encoder that employs token dropping/masking for improved efficiency and enhanced performance in encoding and generation compared to existing works.

Zohar Consulting Services *Mar 2023 - Jun 2024*

President

- Benchmarked compute and inference latencies of LLMs and LMMs on custom hardware.
- Advised on multiple machine-learning pipelines, including cell video classification with SIFAR and conventional detection/segmentation approaches.
- Helped to interview Machine Learning Engineer candidates. Provided guidance and advised on grant proposals to secure funding for research and development projects. Helped secure an NIH grant.

proteanTecs LTD *Sep 2020 - Jan 2024*

Machine Learning & Algorithms Engineer

- Automated data analytics with machine learning, boosting silicon manufacturing yield by 10% and predicting equipment failures with 95% accuracy.
- Led the development of an algorithmic system that automates data analytics tasks composed of (sequential) parametric estimation, outlier detection, and alert collection and analysis for analytic insights.

EDUCATION

Stanford University

Doctor of Philosophy & Knight-Hennessy Scholar *September 2021 - Present*
Electrical Engineering

Master of Science - Computer Science *September 2022 - Present*

Technion - Israel Institute of Technology

Master of Engineering *March 2019 - March 2021*
Electrical & Computer Engineering - Graduated Summa Cum Laude (GPA: 98.4/100)

SELECTED PUBLICATIONS

- **Zohar, O.**, Wang, X., Dubois, Y., Mehta, N. Xiao, T., Hansen-Estruch, P., Yu, L., Wang, X., Felix, J., Zhang, N., Yeung-Levy, S., Xia, X. (2024). Apollo: An Exploration of Video Understanding in Large Multimodal Models. arXiv:2412.10360
- **Zohar, O.**, Farré, M., Marafioti, A., Noyan, M. Xiao, T., Cuenca, P., Zakka, C., Lochner, J. (2025). SmolVLM2: Bringing Video Understanding to Every Device. https://huggingface.co/blog/smolvlm2
- **Zohar, O.**, Wang, X., Bitton, Y., Szpektor, I. Yeung, S., (2025). Video-STaR: Self-Training Enables Video Instruction Tuning with Any Supervision. *Accepted to ICLR 2025.* arXiv:2407.06189
- Hansen-Estruch, P., Yan, D., Chung, C., **Zohar, O.** Wang, J., Hou, T., Xu, T., Vishwanath, S., Vajda, P. Chen, X. (2024) Learnings from Scaling Visual Tokenizers for Reconstruction and Generation. arXiv:2501.09755
- Wang*, X., Zhang*, Y., **Zohar, O.**, Yeung, S., (2024). VideoAgent: Long-form Video Understanding with Large Language Model as Agent. *ECCV2024.* arXiv: 2403.10517.
- **Zohar, O.**, Lozano, A. Goel, S., Yeung, S., Wang, K., (2024). Open World Object Detection in the Era of Foundation Models. arXiv:2312.05745.
- **Zohar, O.**, Huang, M., Wang, K., Yeung, S., (2023). LOVM: Language-Only Vision Model Selection. *NeurIPS (D&B) 2023.*
- **Zohar, O.**, Wang, K., Yeung, S., (2023). PROB: Probabilistic Objectness for Open World Object Detection. *CVPR 2023.*
- Goodman, E. D. *et al.*, (2023). Analyzing Surgical Techniques in Diverse Open-Surgical Videos with Multi-Task Machine Learning. *JAMA surgery.* `doi:10.1001/jamasurg.2023.6262`.
- Zhao, J., Winetraub, *et al.*, (2020). Angular Compounding for Speckle Reduction in Optical Coherence Tomography using Geometric Image Registration Algorithm and Digital Focusing. *Scientific Reports.*

AWARDS, GRANTS & PATENTS

100K Google-HAI research grant *2024-25*
200K HAI-AIMI research grant *2023-25*
Knight-Hennessy Scholar *2021-24*
Intuitive Surgical Best Poster at the SCIEN Industry Affiliates Meeting *Spring 2021*
Patent: "A multifunctional and water-resistant electronic skin empowered with an autonomic self-repair mechanism." *Summer 2021*
The Norman and Barbara Seiden family prize *Spring 2018*
Technion president's award (7x, top 3% GPA) & Technion dean's award (1x, top 15% GPA). *2015-19*

ADDITIONAL EXPERIENCE

LNBD, Technion - Israel Institute of Technology *Nov 2018 - Mar 2021*
Junior Researcher - Soft Electronics

Helped develop state-of-the-art self-healing multifunctional-multilayer electronic skins and sensors.

QUAD Lab, Technion - Israel Institute of Technology *May 2017 - Oct 2019*
Student Research Projects (A & B)

- Developed high-TC Superconducting Nanowire Single-Photon Detectors. Initiated the Selective Growth method, which is currently used to produce YBCO SNSPDs.
- Physical, electrical, and thermal modeling of superconductor-semiconductor tunnel junctions.

OTHER SKILLS

Programming Languages	Pytorch, Python, Julia, *C*
Software & Tools	Working with remote Linux/Vertica/S3/GPC servers, GitHub
Other	LaTeX, MS Office