Alexander Vilesov

Webpage: asvilesov.github.io

Email : vilesov@ucla.edu
Phone : +1-626-390-7241

Education

- **University of California, Los Angeles** — Los Angeles, CA
  PhD in Electrical and Computer Engineering; GPA: 3.97/4.0 — *Sep 2021 - Present (Expected Jun 2026)*
- **University of California, Los Angeles** — Los Angeles, CA
  Master's in Electrical and Computer Engineering; GPA: 4.0/4.0 — *Sep 2021 - Jun 2023*
- **University of Southern California** — Los Angeles, CA
  Bachelor's in Electrical Engineering; GPA: 3.93/4.0 — *Aug 2017 - Jun 2021*

Selected Projects

- **VLM4D: Benchmarking Spatiotemporal Reasoning in Vision-Language Models**:
  - Proposed **VLM4D**, the first benchmark for evaluating spatiotemporal reasoning in Vision-Language Models (VLMs)
  - Created a curated dataset of real and synthetic videos with QA pairs focused on motion, rotation, and perspective shifts
  - Identified core limitations in state-of-the-art VLMs, including failure to integrate visual cues and temporal coherence
  - Demonstrated future improvements can be obtained from spatial grounding or spatiotemporal fine-tuning methods
  - Positioned VLM4D as a foundation for future research on spatial intelligence.
- **Chimera: Creating Digitally Signed Fake Photos by Fooling Image Recapture and Deepfake Detectors**:
  - Developed **Chimera**, the first end-to-end attack that fools both deepfake and image recapture detectors simultaneously while still allowing an image to be cryptographically signed by providence methods such as C2PA.
  - Designed a **hardware-aware two-step GAN-based pipeline** that compensates for physical recapture artifacts (e.g., Moiré) through camera/display-specific tuning. Thus, demonstrating an over **50% drop in detection accuracy** and a 15% increase in attack success rate across state-of-the-art detectors in real-world setups
  - Highlighted a critical vulnerability in future cryptographic photo authentication systems
  - Accepted for publication at **USENIX Security 2025**.
- **CG3D: Compositional Generation for Text-to-3D via Gaussian Splatting**:
  - Developed a novel method for generating scalable 3D scenes from text prompts, leveraging **explicit Gaussian radiance fields** for semantically and physically accurate compositions
  - Addressed key challenges in text-to-3D: **multiscale compositionality, object separability, and physical realism**, surpassing prior methods that struggled with multi-object and scene-level prompts
  - Introduced a method of utilizing **score-distillation sampling** to optimize object composition parameters such as scale, translation, and rotation without requiring bounding boxes
  - Created a framework that enables **flexible scene editing and recomposition**, facilitating object-level transformations from only textual input.
- **EquiPleth: Camera-Radar Fusion for Equitable and Improved Heart Rate Estimation**:
  - Proposed the first remote multimodal camera and radar fusion heart rate estimation method that showed **performance and equity gains of 37% and 70% respectively** over the prevalent camera-based method
  - Theoretically showed the **light transport bias in the camera-based modality** against darker skin tones
  - Designed an **adversarial learning-based pipeline** for fair multimodal fusion
  - Established the **largest multimodal video+radar heart rate monitoring dataset**, with **91 participants**
  - Accepted for publication at **ACM SIGGRAPH 2022**. *Media: UCLA Newsroom*, *Daily Bruin*

Skills Summary

- **Programming Languages:** Python, C++, MATLAB, Bash/Shell
- **AI & Computer Vision:** Large Language Models (LLMs), Vision-Language Models (VLMs), Diffusion Models, Radiance Fields (NeRF/3DGS), fine-tuning (LLMs/VLMs), Deep Learning, Generative AI, Agentic AI systems
- **Frameworks & Libraries:** PyTorch, Hugging Face (Transformers, PEFT, Accelerate, TRL), LangGraph, Diffusers, OpenCV, camera and sensor hardware (ISP), prototyping/3D printing.

Professional Experience

- **University of California Los Angeles** — Los Angeles, CA
  Graduate Student Researcher, Visual Machines Group — *Aug 2019 - Present*
  - PhD student working with **Prof. Achuta Kadambi** (ECE/CS) focusing on equipping **vision-language models (VLMs)** with spatial intelligence — enabling them to understand and reason about **3D structure, physical relationships**, and **spatial context** from multimodal input.
  - Broader research interests include **large language models**, **diffusion models**, **digital health**, and **camera security**. Prior work includes developing **generative representations**, **remote vital sign monitoring**, and **equitable sensing techniques** for clinical environments as evidenced in projects and papers sections.
  - Mentored and led **10+ undergraduate and Master's students** in research and technical development.

- **NASA Jet Propulsion Laboratory** — Pasadena, CA
  Software Engineering Intern — *June 2020 - August 2021*
  - Developed C++ library for on-board processing in Cosmic-2 satellites to acquire in orbit GPS signals of the European Galileo constellation. Applied 2D doppler frequency shift search of GPS codes on raw antenna data to account for high relative orbiting velocities.
  - Developed C++ library for decoding Galileo GPS signals to form Ephemeris and Almanac data and predicting future orbital trajectories of Galileo satellites.

- **University of Southern California** — Los Angeles, CA
  Undergraduate Student Researcher, S2L2 Lab — *June 2019 - July 2021*
  - Worked with Prof. Rahul Jain (Dept. of Electrical & Comp. Engineering), on reinforcement learning algorithms.
  - Designed inverse reinforcement learning algorithms for agents and evaluated reinforcement learning algorithms like Deep Q Learning for image based environments.

Publications and Manuscripts

- Zhou, S.*, **Vilesov, A.***, He, X.*, Wan, Z., Zhang, S., Nagachandra, A., Chang, D., Chen, D., Wang, X. E., Kadambi, A. (2025). **VLM4D: Towards Spatiotemporal Awareness in Vision Language Models**. In Submission. Link
- Chari, P., Can, S. E., **Vilesov, A.**, Chen, H., Srivastava, N., Kadambi, A. (2024). **SteadyDepth: Fast Stable Uncertainty-driven Monocular Video Depth Estimation**. In Submission.
- Park, S.*, **Vilesov, A.***, Zhang, J., Khalili, H., Tian, Y., Kadambi, A., Sehatbakhsh, N. (2025). **Chimera: Creating Digitally Signed Fake Photos by Fooling Image Recapture and Deepfake Detectors** USENIX Security 2025. Link
- Chari, P.*, Harish, A.*, Armouti, A., **Vilesov, A.**, Sarda, S., Jalilian, L., Kadambi, A. (2024). **Implicit Neural Models to Extract Heart Rate from Video**. European Conference on Computer Vision (ECCV). Link
- Del Regno, K., **Vilesov, A.**, Armouti, A., Harish, A. B., Can, S. E., Kita, A., & Kadambi, A. (2024). **Thermal Imaging and Radar for Remote Sleep Monitoring of Breathing and Apnea.** arXiv preprint arXiv:2407.11936. Link
- **Vilesov, A.**, Chari, P., & Kadambi, A. (2023). **CG3D: Compositional Generation for Text-to-3D via Gaussian Splatting**. arXiv preprint arXiv:2311.17907. Link
- Zhao, E., **Vilesov, A.**, Athreya, S., Chari, P., Merlos, J., Millett, K., St Cyr, N., Jalilian, L., Kadambi, A. (2023). **Making Thermal Imaging More Equitable and Accurate: Resolving Solar Loading Biases**. arXiv preprint arXiv:2304.08832. Link
- **Vilesov, A.***, Chari, P.*, Armouti, A.*, Harish, A., Kulkarni, K., Deoghare, A., Jalilian, L., & Kadambi, A. (2022). **Blending Camera and 77 Ghz Radar Sensing for Equitable, Robust Plethysmography**. In ACM Transactions on Graphics (SIGGRAPH). Link

Patents

- Kadambi A., Yuan T., **Vilesov A.**, Nader S. (2024). **Securing Camera and Photography Systems From Deepfakes by Verifying Provenance and Reducing Attack Surfaces**. Patent pending.
- Kadambi A., Kita A., **Vilesov A.**, Del Regno K., Can E., Jalilian L., Harish A., Armouti A. (2024). **Methods and Apparatus to Detect and Classify Forms of Sleep Apnea**. Patent pending.

Honors and Awards

- Awarded the **Artificial Intelligence Rising Scholar Research Award (AIRS)** for the year 2021-2022
- Awarded the **NSF Graduate Research Fellowship Honorable Mention** in 2022
- Nominated for US Congress **DARPA Project Demo** in 2023 for work in multimodal fusion of camera and radar for remote heart-rate detection.

Teaching Experience

- **Computational Imaging and Generative AI (UCLA, ECE239AS)**: Winter 2025
- **Signals and Systems (UCLA, ECE102)**: Spring 2024
- **Honors Digital Signal Processing (UCLA, ECE189)**: Winter 2023
- **Digital Signal Processing (UCLA, ECE113)**: Fall 2022, Winter 2023.

References

- UCLA CS/ECE Professor Achuta Kadambi: achuta@ee.ucla.edu
- UCLA School of Medicine Doctor Laleh Jalilian: ljalilian@mednet.ucla.edu
- NASA JPL Engineer Thomas Meehan: thomas.k.meehan@jpl.nasa.gov