Jurijs (Yuri) Nazarovs

As a senior expert in Computer Vision, specializing in Multimodality Understanding and Vision-Language models, I lead AI research and development across computer vision and foundation models. I support and guide my team members, fostering a collaborative environment. I also acquired an extensive experience in Generative AI during my PhD.

Address: San Jose, CA 95117
Phone: (919) 396-1252
E-mail: **ynaz93@gmail.com**
Website: **jurijsnazarovs.github.io** (**Google Scholar**)

INDUSTRY EXPERIENCE

Amazon, Applied Scientist since August 2024

◇ **Video VLM**: Fine-tuned an Alexa video VLM for per-person grounded activity recognition; Florence-2-style task-specific tokens cut inference-time token cost on captioning and grounding heads by ∼**60%**. Mentored a junior scientist through a peer-reviewed conference submission.

◇ **Text-to-Text model, Alexa AI**: Trained an LLM for context-aware response generation (actionable insights, user affinities, in-the-moment outputs) with task-specific-token fine-tuning; replaced brittle n-gram metrics (BLEU/ROUGE) with a product-aligned **LLM-as-judge** evaluation pipeline.

◇ **Multi-camera Scene Understanding**: Architected a multi-camera VLM pipeline aggregating synchronized home feeds to overcome occlusion and cross-room context loss, lifting recall ∼**20%** over single-camera baseline. Benchmarked **Qwen3-Omni**, **Nova Lite**, and **Gemini** across input modes (video, sequential/interleaved frames) for cross-camera person identification and behavior analysis (meetings, mood, photo moments).

◇ **Person Re-identification**: Designing a tiered cache gallery for scalable, low-latency re-identification across distributed video streams; current system lifts recall ∼**40%** over a face-detection-only baseline.

Ambient.ai, Senior Applied Research Scientist April 2023 - August 2024

◇ **Multimodality**: led a natural language video search project, focusing on deploying the ImageTagging model on devices, utilizing quantization with fine-tuning for enhanced performance. To manage hardware limitations for frame rate, developed a novel dynamic weighted frames sampling method to focus on frames of interest.

◇ Implemented a **zero-shot** object detection and segmentation pipeline using advanced Vision Language Models, Grounding DINO and SAM, optimized through deployment of an Efficient Vit.

◇ Directed a project on **Incremental Learning** for Object Detection models, utilizing distillation and creating an automated labeling pipeline that leveraged the Grounding DINO model, reducing data annotation costs in two times.

EDUCATION

University of Wisconsin - Madison, Madison, WI. PhD, Statistics
University of Wisconsin - Madison, Madison, WI. MS, Computer Science
Duke University, Durham, NC. MA, Economics

SKILLS

Computer Vision, Foundation Models, Multimodality, VLM, LLM, Video Understanding, Grounded Activity Recognition, Person Re-identification, Scene Understanding, Synthetic Data Generation, Fine-tuning (task-specific tokens, distillation, quantization), Evaluation Pipelines (embedding- and LLM-based metrics), Generative Models, Deep/Machine Learning, Trajectory Prediction, Perception, Probabilistic Models, VAE, BNN, GAN, Python (PyTorch, TensorFlow), R, Bash, Linux, AWS

PUBLICATIONS

Grounded Human-Attributed Description and Activity Recognition in Videos (GHADAR), ECCV 2026 (under review). Introduced the GHADAR task for per-person, open-set attribute and activity description in multi-person videos, along with **AVA-Captions** – the first large-scale grounded dataset of this kind, extending AVA-Actions via VLM-generated captions and identity-aware deduplication. Proposed **CAMP** (Constrained Attention Masking-based Pretraining), a two-stage VLM training strategy that explicitly leverages grounding through attention-mask constraints, outperforming SOTA VLMs; also introduced a VLM-driven evaluation framework comparing video and prediction at the concept level rather than via n-gram/embedding metrics.

Image2Gif: Generating Continuous Realistic Animations with Warping NODEs, CVPR 2022 (AI4CC workshop). Introduced a novel Deep learning Module, Warping Neural ODE, as a Video Frame Interpolation (VFI) mechanism, to generate GIF between two conceptually far apart frames. Method allows to generate unlimited number of FPS, making smooth VFI.

Mixed Effects Neural ODE: A variational approximation for analyzing the dynamics of panel data, UAI 2021 (26% acceptance rate). Introduced the temporal generative model, Mixed-Effect Neural ODE, which allows to model uncertainty

like SDE, but use ODE solvers in combination with DNN, for trajectory prediction of physical processes, humanoids and reconstruction of 3D brain scans of Alzheimer's disease progression.

Functional NODE - sampling of trajectories. Introduced a new Functional NODE framework which allows to sample trajectories in a VAE-like procedure, e.g. human/skeleton actions, physical processes, and other and perform statistical inference.

Understanding Uncertainty Maps in Vision with Statistical Testing, CVPR 2022 (25% acceptance rate). Introduced a stable diffusion like model, Warping Neural ODE combining with Random Fields theory, to derive significant regions of the Uncertainty Maps obtained from probabilistic DNN (BNN/VAE) in image generation and perception settings, like segmentation.

Improving Robustness of VQA Models by Adversarial and Mixup Augmentation. Introduced an adversarial objective function to train the VQA (VLM), based on UNITER-like architecture with BERT component, to improve VQA robustness to linguistic variations and visual manipulations.

Variational Sampling of Temporal Trajectories. New method for trajectories synthesis.

Graph Reparameterization for enabling 1000+ Monte Carlo Iterations in Bayesian Deep Neural Networks, UAI 2021 (26% acceptance rate). Developed a new framework to construct an MC estimator for the KL term, which significantly decreases GPU memory needed to run VI version of Bayesian Neural networks and improves runtime. Memory savings allow us to run up to 1000 or more MC iterations on a single GPU.

Radial Spike and Slab Bayesian Neural Networks for Sparse Data in Ransomware Attacks, U.S. Patent.

Ordinal Quadruplet: Retrieval of Missing Labels in Ordinal Time Series, U.S. Patent.