Negar Nejatishahidin

571-585-8143 | negarnejatiuni@gmail.com | N-NEJATISHAHIDIN.github.io | linkedin.com/in/negar-nejati

Summary

I hold a Ph.D. in Computer Science with a specialization in computer vision and deep learning. My research focuses on enhancing scene understanding and interaction through innovative perception systems, generative multimodal architectures, and large language models.

Education

George Mason University, **Fall, 2019 – Spring, 2025**

Doctor of Philosophy in Computer Science (Supervised by Dr. Jana Kosecka), *GPA : 3.8/4*

Technical Skills & Research Interests

Programming Languages: Python, C/C#/C++, Java, MATLAB, SQL.
Technologies & Frameworks: PyTorch, TensorFlow, Hugging Face, Unsluth, Ollama, llama.cpp, PyTorch3D, OpenCV, Open3D, COLMAP, OpenDroneMap, OpenSfM, Scikit-learn, SciPy, Pandas, Keras, NumPy, Docker, Kubernetes, LangChain, FastAPI, AWS, Google Cloud Platform (GCP).
Machine Learning & Deep Learning Models: Transformers, CNNs (Convolutional Neural Networks), RNNs (Recurrent Neural Networks), GNNs (Graph Neural Networks); Phi, Llama, Qwen, Mistral AI, LLava, CLIP, ALIGN, DINO, DETIC, DETER, Grounding DINO, SAM (Segment Anything Model); Stable Diffusion, Diffusion Models; Ensemble Learning, Unsupervised Learning, Self-Supervised Learning, Supervised Learning, Contrastive Learning, Representation Learning; Pretraining, Finetuning, Few-Shot Learning, Meta-Learning, Prompt Engineering, Transfer Learning, Domain Adaptation.

Professional Experience

HP Inc. **Dec. 2024 – Present**

Senior Computer Vision Engineer

- Fine-tuned specialized Vision–Language Models (VLMs), Small Language Models (SLMs), and Large Language Models (LLMs) for efficient on-device deployment, optimizing both performance and memory footprint.
- Developed an internal evaluation framework to benchmark state-of-the-art VLMs and LLMs, enabling standardized performance analysis across projects.
- Built synthetic data-generation pipelines for generative AI, leveraging external state-of-the-art models to create custom datasets.

Humane **May 2023 – Aug. 2023**

Computer Vision Research Intern

- Enhanced hand-tracking pipeline generalization by integrating semi-supervised learning techniques, enabling the incorporation of large-scale unlabeled data into the training process. Improved model accuracy by 13% based on internal evaluation.
- Designed and automated dataset creation, collection, and annotation processes for large-scale evaluation of machine learning applications.

Zillow Group **May 2022 – Dec. 2022**

Computer Vision Research Intern

- Proposed the first end-to-end architecture for global camera localization, delivering an absolute improvement in rotation accuracy and comparable translation performance relative to the state of the art.
- Supervised by Dr. Sing Bing Kang.

Patent

Automated Inter-Image Analysis of Multiple Building Images for Building Information Determination

- US Patent Application 18/114,951
- W.A. Hutchcroft, Y. Li, M. Narayana, **N. Nejatishahidin**

Publications

Graph-CoVis: GNN-based Multi-view Panorama Global Pose Estimation (CVPRW 2023)

- OmniCV workshop at IEEE / CVF Computer Vision and Pattern Recognition Conference (CVPR).
- **Best paper award.**
- ***N. Nejatishahidin****, Zillow Group*

Uncertainty-Aware Open-Vocabulary Zero-Shot 3D Semantic Segmentation

- To be submitted.at WACV 2025
- ***N. Nejatishahidin****, J. Kosecka*

Structured Spatial Reasoning with Open Vocabulary Object Detectors (ICRA 2024)

- ICRA 2024 Workshop, Back to the Future: Robot Learning Going Probabilistic
- An extenuation Under-review at ICRA 2025
- ***N. Nejatishahidin****, J. Kosecka*

Object Pose Estimation using Mid-level Visual Representations (IROS 2022)

- IEEE/RSJ International Conference on Intelligent Robots and Systems.
- ***N. Nejatishahidin****, P. Fayyazsanavi, J. Kosecka*

Fingerspelling PoseNet:Enhancing Fingerspelling Translation with Pose-Based Transformer Models (WACV 2024)

- WVLL workshop at IEEE/CVF Winter Conference on Applications of Computer Vision (WACV).
- *P. Fayyazsanavi,* ***N. Nejatishahidin****, J. Kosecka*

Projects

Multi-modal 3D Scene Semantic Segmentation **Jan. 2024 – Present**

- Proposing a novel approach for Open vocabulary 3D semantic segmentation for unseen environment from un-posed RGB images.
- Developing 3D representations such as point cloud, voxel, textured mesh, and NeRF 3D representations with the aid of structure from motion pipelines like Colmap, OpenSFM, and OpenDronMap.
- Incorporating state-of-the-art monocular depth estimation models such as ZoeDepth, PatchFusion, and Marigold to improve reconstruction quality and address noisy depth data.
- Proposing a pipeline based on Vision-Language Models (VLMs) such as CLIP, ALIGN, RAM, GroundingDINO, GPT-4, and SAM to develop a robust and generalizable 3D semantic segmentation model that can be queried using natural language. By integrating these advanced VLMs, the pipeline enables more intuitive and flexible interaction with 3D data.
- Enhancing semantic segmentation through uncertainty estimation using an ensemble model.

Spatial Reasoning using 3D geometric priors **Aug. 2023 – Sep. 2024**

- Bench-marking the Generative models (Stable Diffusion, DALL·E, Glide, ControlNet, Imagen, Latent Diffusion Models (LDM)) for the task of 3D spatial reasoning Using prompt engineering on LLAVA, LLAVA-Next, and Grounding Dino.
- We propose a Spatial Relation Detection model built on top of state-of-the-art open-vocabulary object detectors, such as Grounding DINO, DETIC, and Detectron2, enriched with 3D geometric features derived from masked 3D object point clouds. It outperforms LLAVA in spatial relation detection by more than 40%.
- Developing an auto-labeling approach for spatial reasoning, applicable to internet-scale data.

Camera Localization **May. 2022 – Jan. 2023**

- Proposing the first end-to-end architecture for multi-view panorama global pose estimation.
- Proposing the first global pose estimation architecture that can effectively handle varying numbers of input panoramas.
- Developed a novel Graph Neural Network architecture, with Transformers as the Node and Edge components.The message passing scheme that enables the GNN to leverage deep representations of dense visual overlap and boundary correspondence constraints, to better estimate global pose.

Object Pose Estimation **May. 2020 – Sep. 2021**

- Developed object pose estimation model on the top of generic mid-level representations (normals, reshading, and depth).
- Introduced 6D pose annotations for the Active Vision Dataset, labeling 6,337 object poses and 3D bounding boxes.
- Achieved 35% performance improvement in low training data regime over SOTA with only 25% of the training data

Pointcloud completion **Oct. 2021 – May. 2022**

- Complete partial point cloud with Transformers, reducing generation error by over 3% with only the encoder.
- Introduced novel positional-aware embeddings leveraging PointNet++, replacing conventional learnable embeddings.

Depth Completion **Nov. 2020 – May. 2021**

- Designed a method to use surface normal, semantics, and geometric information for depth completion.