Sina Malakouti

siinamalakouti@gmail.com or sem238@pitt.edu | sinamalakouti.github.io | sinamalakouti | sinamalakouti

Legal status in the US: Permanent Resident (Green Card holder)

Education

University of Pittsburgh — Aug 2020 – Jan 2026
PhD in Computer Science — *Pittsburgh, PA*

- Advisor: Adriana Kovashka
- Committee: Adriana Kovashka, Boqing Gong, Xiange Lorraine Li, Milos Hauskrecht
- Thesis: *Compositional Gaps in Object Representations for Generative and Discriminative Models*

Amirkabir University of Technology — Sep 2015 – May 2020
B.Sc. in Software Engineering — *Tehran, Iran*

Selected Peer-Reviewed Publications

- **Role Bias in Text-to-Image Diffusion Models: Diagnosing and Mitigating Compositional Failures through Intermediate Decomposition**. **S. Malakouti**, A. Kovashka. **Under Review**
- **Benchmarking VLMs' Reasoning About Persuasive Atypical Images**. **S. Malakouti***, A. Aghazadeh*, A. Khandelwal, A. Kovashka. **WACV 2025**
- **Incorporating Geo-Diverse Knowledge into Prompting for Increased Geographical Robustness in Object Recognition**. K. Buettner, **S. Malakouti** *(major contributor)* , X.L. Li, A. Kovashka. **CVPR 2024**
- **Semi-Supervised Domain Generalization for Object Detection via Language-Guided Feature Alignment**. **S. Malakouti**, A. Kovashka. **BMVC 2023**
- **DeepTreeNetworks: A New Symbolic Deep Architecture**. **S. Malakouti***, Z. Ahmadi*, S. Kramer. **DeCoDeML Workshop, ECML PKDD 2019**

Experience

Graduate Research Assistant — Aug 2020 - Present
University of Pittsburgh — *Pittsburgh, PA*

- Researching on making ML methods more **robust** and capable of understanding and **reasoning** about complex and **compositional scenes**, focusing on **vision-language foundational models** (VLMs and text-to-image generative models). Published papers in popular computer vision venues (CVPR, WACV, BMVC).
- Benchmarking T2I diffusion models across culturally grounded objects, activities, rituals, and diverse scenes.
- Designing a modular framework combining **compositional generation** (via intermediate decomposition) with **descriptor-guided control** using a Q-former adapter for culturally grounded object and scene generation.
- Proposed **RoleBench**, a benchmark for evaluating **directional role generalization** in action-based T2I generation (e.g., "mouse chasing cat"), revealing systematic **role collapse** in SOTA models.
- Developed a lightweight compositional framework using **LLM-guided intermediate decomposition**, improving rare composition generation with **15.2-point bias reduction** and **>70% human preference**.
- Curated **PersuasiveAdVLM**, the first benchmark for textbfcompositional and multi-step reasoning in MLLMs on persuasive ads with **atypical/unusual** object compositions; revealed **visual reasoning gaps** between MLLMs and LLMs and **lack of visual reasoning** in MLLMs.
- Proposed an **atypicality-aware chain-of-thought** method, improving zero-shot reasoning by 40% on abstract and unusual visual relations.
- Developed a **novel soft prompt learning** method for CLIP by distilling LLM knowledge, achieving sota performance in cross-cultural object recognition benchmarks.
- Developed a **multi-scale contrastive-based** method that learns robust visual features by preserving semantics in the language space, improving cross-domain object detection by 12% without the need for target domain data.

Applied Scientist Intern — May 2024 - Sep 2024
Prime Video, Amazon — *New York, NY*

- Led research on multimodal content understanding & duplicate detection. Developed a novel CLIP-based fusion model and multimodal chain-of-thought (Claud, InternVL), achieving > 10% improvement. **S3, SageMaker**

Applied Research Intern — May 2023 - Aug 2023
Search Science, eBay — *San Jose, CA*

· Employed **vision-language models (CLIP)** and a novel **transformer-based Mixture-of-Modality-Experts fusion** model, significantly boosting results on search and ranking tasks. **PyTorch, Spark, Hadoop**

Computer Vision Intern May 2022 - September 2022
Apple *Cupertino, CA*

· Developed efficient models for computer vision and Image Processing tasks, achieving enhanced performance and efficiency over state-of-the-art methods and baselines. **Python**, **PyTorch**, and **Matlab**.

Machine Learning Research Assistant, Intern July 2018 - Sep 2019
Johannes Gutenberg University *Mainz, Germany*

· Designed an efficient symbolic deep network using differentiable decision trees, effective on imbalanced data.

Technical Skills

Programming Languages	Python, Java, MATLAB, SQL, C, R, JavaScript, HTML/CSS
AI & CV Methods	CNNs, RNNs, Transformers & Attention Mechanism, Vision-Language Models (VLMS), Large Language Models (LLMs), Multimodal LLMs (MLLMs), Text-to-Image (T2I) Diffusions, Contrastive Learning, Semi-Supervised Learning, Domain Adaptation/Generalization (e.g., Pseudo Labeling, KD, Student-Teacher, Consistency), Parameter-Efficient Learning (e.g., Soft Prompting, Adapters, LoRa), Multiple Instance Learning (MIL), Alignment (RLHF, DPO)
ML Tools	PyTorch, TensorFlow, Keras, Scikit-learn, DL4j, Weka, Numpy, Pandas
Big Data & Databases	Hadoop, Spark, S3, MySQL, MongoDB, SQLite
Other	Data Engineering, Object-Oriented Design, MVC, Problem-Solving

Other Projects

- **Weakly Supervised Object Detectors Robustness Toward Domain Shift**
  Python, PyTorch, Weakly-Supervised Object Detection (WSOD), Domain Robustness

· Hypothesized stronger reliance of WSOD on domain-specific features than fully supervised methods. Improved detection on unseen domains by 2% using consistency regularization with style transfer.

- **Multimodal Transformer Fusion For Depression Prediction**

· Developed a novel approach for depression severity prediction by creating a joint representation of unaligned video, language, and audio features through **multimodal transformer fusion**.

- **MuST for Semi-Supervised Medical Image segmentation**
  Python, PyTorch, Data Augmentation, Consistency Regularization, Semantic Segmentation

· Proposed a **feature-space augmentation consistency** approach for brain lesion segmentation, achieving sota performance with only 3% labeled data

- **Image-Caption Discourse Coherence Relation Prediction**
  Python, PyTorch, Self-Supervised Learning, Discourse Relation

· Enhanced semi-supervised image-text discourse-relation prediction in a semi-supervised manner by enhancing self-supervised models (e.g., ViLBERT, SwAV) by employing a self-training-based approach.

Professional Services

Conference Reviewer: CVPR, ICCV, ECCV, NeurIPS, AAAI, EMNLP, WACV
Co-Organizer: Demographic Diversity in Computer Vision Workshop, CVPR 2025

Honors & Awards

· **Doctoral Consortium**, Winter Conference on Applications of Computer Vision (WACV), 2025
· **Outstanding Reviewer Award**, European Conference on Computer Vision (ECCV), 2024
· **Travel Award**, Department of Computer Science University of Pittsburgh (2023)
· **Full SCI Fellowship**, University of Pittsburgh (2020)
· **Honored as an outstanding student**, Amirkabir University of Technology (2015-2020)

Extra Curricular & Leadership

President of Student Scientific Chapter Jan 2017 - March 2018
Computer Engineering, Amirkabir University of Technology *Tehran, Iran*

· Organized 70+ national and international contests, talks, and workshops in collaboration with Technische Universität München, Germany, and KTH Royal Institute of Technology, Sweden.