Research Statement

Milad Rabiei **PhD Applicant**

"*The real power of artificial intelligence and robotics comes when it's applied to things we don't do well*."
- Rodney Brooks

I am intrigued by modern sciences, and I have planned to dedicate my time developing intelligent systems. My unwavering commitment to advancing this field, my experiences, and my dedication to mentorship and education have prepared me for this next chapter in my academic journey.

During my undergraduate years, I had the privilege of establishing the Vision section at SBU's Robotics and Intelligent Automation Lab under the guidance of Dr. Mohammad Hossein Moaiyeri.

One of my pivotal contributions was the design and implementation of a novel landmark tracking method, enhancing road lane following accuracy for autonomous vehicles. Following the basic principles of density-based clustering, yet alternating some requirements relative to our domain space, I came up with an area-oriented unsupervised segmentation algorithm that ignores possible inevitable noise and is not computationally expensive for our problem, capturing and classifying road boundaries. This algorithm addresses the limitations of traditional sliding windows technique; By tracking the landmarks through any angle, it achieves robustness even through very sharp turns, following any group of points almost as good as DBSCAN clustering without its almost huge computational overhead.

Additionally, I tackled the inadequacy of object detection datasets for our robot by designing and implementing a synthetic image generation and augmentation pipeline. Initially experimenting with YOLOv3 and progressing to more advanced models, I tested multiple datasets like GTSDB and Mapillary. Due to our specific set of signs, instead of tediously labeling an entire new dataset by hand, I decided to create one by systematically overlaying our specific objects on various close-enough backgrounds, performing image augmentations if needed, and saving the labels. The model trained on this synthetic dataset outperformed the one trained on the scant manually-labeled dataset across various image conditions. The image generation tool I developed remains in use by my peers and younger colleagues for specific object detection tasks.

My efforts significantly contributed to our team's victories at the RoboCup IranOpen, the country's premier international robotics competition. I actively participated in the research, resulting in a manuscript to be published that encapsulated our findings and insights.

To broaden my knowledge, I delved into information theory and developed an interest in generative models. This led to collaborating with Dr. Vahideh Moghtadaiee and Dr. Mina Alishahi on using GANs to generate differentially-private indoor positioning data with both regression and categorized labels. My at-the-time limited knowledge of GANs, prompted me to actively seek out and acquire the requisite expertise. I also delved into an exploration of privacy, particularly within the context of machine learning.

Throughout this ambitious project, I gained extensive knowledge by tackling research challenges in our collaborative sessions. I conducted practical experiments with generative AI, leveraging my skills in data engineering and academic writing to significantly contribute to the paper and the project's success.

The advancements in the combination of information theory and GANs can further help in understanding the connections among different Gen-AI variants and how they have evolved. This experience reinforced my belief in the power of interdisciplinary research.

While my undergraduate projects have followed an exploratory path, my primary aim in my graduate studies is to concentrate on innovative and impactful research in modern computer vision. Tackling real-world challenges of self-driving cars exposed me to several research concentrations.

My previous work in object detection for autonomous vehicles was concluded smoothly because with a proper dataset, recent advancements in computer vision and deep learning dominate the 2D object detection challenge almost perfectly. However, certain considerations remain. Since due to the nature of the problem, intelligent systems could not rely completely on information from a 2D information base, what can be done to reach a similar intelligence measure with 3D visual perception as well as 2D? I am interested to investigate improvements on multi-view systems for 3D object detection task.

Furthermore, ideally it makes sense to expect an existing cyber-physical system to be able to understand how regular 3D objects are formed. By today's standards, such generalization comes from a large dataset with a robust-enough representation. Since such material is not currently publicly accessible, I'm curious to come up with how we can design architectures and experiments to reach such understanding.

Considering the important advantage of object mesh estimation, one of the main challenges for adopting data-driven approaches for shape processing is selecting a proper representation for the 3D data, like polygonal meshes, point clouds, parametric functions, etc. I am curious about coupling encoders with a scoring algorithm focused on learning an intermediate geometric representation.

It is always fun to discover what implicit information a generative artificial model has learned, and furthering my knowledge of machine learning theory, I'm interested to come up with profitable and differentiable data transformations that can be used in modern deep learning models.

Additionally, in the world of autonomous vehicles, datasets are semantically apart from each other, e.g., simulation and real-world scenes or views of different cameras. The domain shift can become expensive, so I wonder if it would be possible to concatenate datasets of video clips and the maneuver control feedback data for a multimodal scenario.

I am also interested in introducing skepticism to our current artificial perceptual understanding. During my paper readings, I got familiar with some architectures that evaluate vision and audio tapes correspondingly, which by itself is fascinating. Yet, I was wondering about how we can improve such evaluations. Is it really enough to be able to acknowledge or decline the correspondence of different signal types?