ANIRUDDHA GANGULY https://aniruddhaganguly.web.app

Statement of Purpose

I was introduced to research in my sophomore year, thanks to my friend and partner in every academic or development project. Since then, it has been my second-best shield against all the perils of life (the first being music). When reading a new paper, learning a new framework, or exploring a new topic - I can block everything else out and focus singularly on it. Over time, research became something I excelled at. That is why I allowed myself to dive deep into academic research during my junior and senior years.

I started with **Computer Vision.** I got into this domain because I was fascinated with Tesla's self-driving vehicles. When I first read the reports on the initial experiments of autopilot for the Tesla Model S in 2016, I was impressed. This fascination was a big reason why, under Dr. Shamsuzzoha Bayzid and Zahid Rahman's supervision, I chose autonomous vehicles as my problem domain. Object detection and localization have made leaps in the past two decades, with state-of-the-art 2D models now regularly achieving astonishingly high accuracies in detection and localization. 3D is a different story, though. Models that use sensor fusion to combine RGB images and LiDAR data to detect and localize objects in a 3D point cloud are improving but are only sometimes fit for real-time performance. Many of these models use 2D object detectors to select a frustum-shaped region within a point cloud and iterate through the points within that segment. It was evident that they iterated over a lot of irrelevant data. This led us to develop **3D-FFS**, a novel approach to make sensor fusion-based 3D object detection networks significantly faster using a class of computationally inexpensive heuristics. We leveraged aggregate intrinsic properties of point cloud data to constrain the 3D search space and eliminate irrelevant data points. We expected this to significantly reduce training time, inference time, and memory consumption without sacrificing accuracy. I integrated *3D-FFS* with *Frustum-ConvNet*, a prominent sensor fusion-based 3D object detection and localization model, and improved training and inference times by up to 62.80% and 58.96%, respectively while *reducing the memory usage by up to 58.53%*. What I enjoyed most was the experience of working with a vast and complex codebase for a state-of-the-art deep-learning model and eventually understanding it enough to make substantial modifications to it. I had done PyTorch courses before, but this was when I learned how to write deep learning code from scratch and integrate modular code into a complex PyTorch codebase. The best part about our project was, contrary to our expectations of encountering at least a noticeable drop in accuracy, our integrated model instead *achieved 0.36%, 0.59%, and 2.19% improvements in accuracy* for the Car, Pedestrian, and Cyclist classes of the KITTI 3D Object Detection Benchmark Dataset. This work led to my **first publication as co-first-author** at the **2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2021)**.

During my undergraduate thesis, I explored a new domain - *Natural Language Processing*. I picked *Bangla Text Summarization* as my undergraduate thesis topic under Dr. Mohammad Eunus Ali. We partnered with a startup that has been running a bite-sized news summary service for a long time. The dataset initially contained over 150 thousand article-summary pairs, generated and reviewed by a team of 15 editors. I manually reviewed and processed the data to generate **RIDMIK^{+}**, a *standardized dataset of 77,336 high-quality article-summary pairs*. I ran intrinsic quality checks on our dataset to find metrics like the n-gram ratio, extractiveness/abstractiveness measure, compression ratio, and redundancy ratio for the dataset. I also ran a few baselines using state-of-the-art models for English text summarization. I modified MatchSum, a BERT-based Siamese network, and incorporated pre-trained embeddings and a custom rouge library for Bangla. MatchSum is trained to determine how semantically similar a candidate summary is to the original article, for which it needs an input module to generate candidate summaries. I used TextRank, a simple Bidirectional LSTM, and SummaRuNNer as the input modules to run experiments. The Bidirectional LSTM version performed commendably on a dry run using our dataset, reaching a rouge-2 score of over 0.6. I am currently running ablation tests, experimenting with various lightweight input modules, and building a dataset website. We are presently evaluating appropriate venues to submit to and parallelly extending the dataset.

The promising results of my first project led me to join another research project under the supervision of Dr. Tanzima Hashem and Dr. Mohammad Eunus Ali. This is a long-term project in partnership with PFDA, a renowned organization in Bangladesh that works for individuals with cognitive disabilities. We started our work with an explorative study on the difficulties faced by individuals with Autism Spectrum Disorder (ASD) during the COVID-19 pandemic, where we looked at how well existing remote learning platforms can accommodate special

needs education. The knowledge we gathered through semi-structured interviews with students and their caregivers was crucial in laying the groundwork for further studies. We determined a number of technical, social, and pedagogical challenges faced by said individuals and their caregivers. Based on these, we proposed design recommendations and ethical considerations for designing a specialized remote learning platform for individuals with cognitive disabilities, particularly those with ASD. We have submitted the manuscript to **CHI 2023**, and it is now **in review**.

For the second task in our work with PFDA, we identified a few core problems faced by individuals with ASD. We discovered that attentiveness classification for people on the spectrum still needed to be explored. Features like *gaze detection* and *activity recognition*, used in the case of neurotypical individuals, are not sufficient in the case of individuals with ASD. We conceptualized *the first deep learning-based attentiveness classifier exclusively dedicated to students with ASD*. Our ensemble contains **five feature extractors** - an atypical movement recognizer, a working status detector, a gaze detector, an activity recognizer, and a base attentiveness feature extractor. We use a TSN for the *atypical movement recognizer* and *gaze detector* since *TSN networks* can utilize spatio-temporal information of the entire video using sparsely sampled snippets. We use a 2D CNN with a TSM for our working status detector. We use a Two-Stream Inflated 3D ConvNet (I3D) for activity type detection, which includes a two-stream architecture combining RGB image sequences and optical flows to detect temporal features. Finally, we have a separate SlowFast module to determine intermediate raw features of attentiveness classification. Additionally, due to the absence of prior work on this task, we set up three baselines using ResNet-101, ResNet-50 + LSTM, and Inception-V3 + LSTM networks. Our proposed architecture outperforms all of these baselines substantially. Besides exploring PyTorch in great depth during the project, I had to draft the manuscript with little assistance from my supervisors. Our work has been **accepted and presented** at the **10th International Conference On Affective Computing & Intelligent Interaction (ACII 2022)**.

My research interests include further exploring 3D object detection, 3D semantic segmentation, 3D reconstruction, few-shot image classification, tumor segmentation works, and action detection tasks, along with their downstream applications. I have always had a fascination for problems within the medical domain. This led to my current independent project, which deals with detecting the likelihood of lung carcinoma from 3D point clouds of the lungs constructed from High-Resolution CT Scans. That is why I am parallelly attempting to build a DICOM dataset of HRCT scans of lung carcinoma. Dr. Debashish Ganguly, a Radiology and Imaging Specialist at LABAID Hospital, is supervising curation and annotation. My current research goal is to explore **3D-focused computer vision tasks**, the **application of vision in robotics**, and **affective computing**. I am also open to exploring i**ntersectional and multi-modal research incorporating NLP into vision**.