

My academic interests mainly lie in research around the intersection of learning algorithms and vision. To this extent, most of my past work has been around this area. Furthermore, I also enjoy working on large-scale software engineering problems. For the rest of this cover letter I share some of my most recent works relevant to the project I'm applying to.

This summer, I am involved in academic Computer Vision research at Qualcomm AI Research hosted by Guillaume Berger and Rolan Memisevic. My primary focus lies in two areas: precise video generation, which is currently under submission, and video question answering, which is still a work in progress. I first-authored a paper on a novel diffusion model designed for precisely guided video generation based on uniquely modifying classifier-free guidance. This paper also introduces a new, large dataset comprising 165,000 videos. I have conducted all the large-scale experiments necessary for this research, which included developing software to manage experiments on more than 200 GPUs.

Most recently at UofT, I was supervised by Prof. David Lindell. The preprint that I worked on was about developing new Diffusion Models that could handle traditional graphics problems and we proposed to pose an entire image signal processing pipeline (with prompts for instance: "What this image would look like in high exposure") as a diffusion model problem. Furthermore, at UofT I also worked on another preprint paper on generating spatial audio for images and videos to accompany the visual content. I also recently worked on a project with Prof. Pascal Tyrrell and Prof. Rahul Krishnan on (very recently accepted to PMLR) building new NeRF models. We particularly proposed (1) a replacement for ray tracing involving not sampling colors and densities like NeRFs do (2) a new rendering algorithm to replace volumetric rendering, and (3) using a diffusion model trained on occupancy grids to serve as a regularizer.

Last summer, I worked on academic Computer Vision research at Civo Ltd. hosted by Josh Mesout. We developed a new multimodal speech and vision model, I not only led the research idea but also implemented and ran the experiments myself. I also led the development of a machine-learning runtime for our product, "Recite", which increased performance by approximately 120-200% compared to other optimized runtimes.

I also believe I have a good understanding of software for machine learning. In the past, I have also worked very closely with CUDA and Triton, and I have also had one first-author and one second-author (oral) paper at the PyTorch Conference: one of which was about building specialized backprop kernels for NeRF training to potentially train NeRFs on small devices and the other was about building capabilities to AOT-compile NeRFs to WebAssembly code. Furthermore, I am also one of the maintainers of Kubernetes, one of the most popular orchestrators for software at scale where I have contributed to building new APIs in the project related to machine learning capabilities, most recently contributing to "Dynamic Resource Allocation".

I strongly believe that some of my past experiences would be very helpful in undertaking the CSC494 project and work on impactful interesting research.

Regards,
Rishit Dagli