Cover Letter – Samarth Mishra

I am a 5th year PhD student in computer science at Boston University, fortunate to be co-advised by Venkatesh Saligrama and Kate Saenko. My research broadly lies in learning with limited supervision. Prior to joining BU, I was a master's student at Georgia Tech, advised by Jim Rehg, and prior to that, I did my undergrad in computer science at IIT Bombay.

Through my MS and PhD, I have worked on different data constrained recognition problems : continual, few-shot/zero-shot, cross-domain transfer learning, which have been published in conference venues (CVPR, ICCV, NeurIPS, BMVC) and journals (TPAMI). During my PhD I have done research internships at MIT-IBM Watson AI Lab (2021, 2022) and at Amazon (2023). The focus of my PhD thesis is finding high quality synthetic data. These range from simpler procedural generations (such as image transformations) or more expensive data using foundation model text-to-image generators. In prior work, with graphics simulation as a source of synthetic images, I showed it is possible to optimize parameters of synthetic data generation and learn a meta-network as a mapping between a downstream task and these optimal parameters. In ongoing work, I found that a few (≈ 1000) high-quality model-generated synthetic data (synthetic captions generated from LLMs in this case) are sufficient to improve visio-linguistic compositional reasoning of vision language foundation models such as LlaVA.

I believe synthetic generation is a key ingredient to finding good quality data for today's models. Specifically in the context of large multimodal models, high quality data in the post-training stage can be crucial for improving specific model capabilities. Synthetic data from generative models might be high-quality but expensive and their cost-quality tradeoffs might be important to study and characterize. A right mix of low-cost procedural generation and model-generated synthetic data might provide the best tradeoff. Optimization of such generated data is also underexplored but possibly of high practical value especially with fewer synthetic samples. I believe with my experience I am positioned well to conduct work in this direction to drive foundation model capabilities.