Yash Sanjay Bhalgat **Statement of Purpose**

Senior ML Researcher, Qualcomm AI Research Webpage: `https://yashbhalgat.github.io`

My interest in pursuing a PhD in Computer Science is mainly driven by a desire to study the theoretical aspects of Geometric Deep Learning and work on its applications to 3D Computer Vision, particularly focusing on the inference efficiency characteristics of these models. Fundamental differences arise when we move from the classical 2D setting of images to non-Euclidean manifolds such as point clouds and meshes. And I am very excited to explore this area of geometric deep learning which deals with the extension of deep learning techniques to such data. Additionally, rapidly emerging applications in autonomous vehicles and AR/VR systems have led to tremendous progress in such algorithms for 3D perception, but these still fall short of the low-latency and/or low-power requirements of these applications. My long term research goal is to develop algorithms for designing efficient geometric deep learning architectures 3D Vision use-cases with considerations given to speed/computation time as well as other aspects like power/memory footprint of these models.

I am also fascinated by the area of unsupervised/self-supervised representation learning. In my PhD research, I would like to explore self-supervised algorithms that incorporate various inductive biases available in 3D geometries to learn meaningful representations for 3D data or more generally multi-modal data (3D + 2D + various sensors) that can help models generalize to challenging environments.

In my past 2 years at Qualcomm AI Research (QAIR), I have been privileged to work with and learn from amazing researchers from core deep learning, systems/hardware as well as wireless communications. My research has been focused on algorithm and system design to develop efficient deep networks for resource-constrained edge devices. As my first project at Qualcomm, I led a team that won the 3^{rd} position in the MicroNet competition at NeurIPS 2019, where the goal was to create the most efficient model (w.r.t. both model size and inference cost) that can achieve 75% top-1 accuracy on ImageNet. In the 2-months span of this competition, I implemented several model quantization/pruning methods from literature and eventually proposed a novel effective method to quantize deep neural networks to extremely low bit-widths. Our low-bit quantization work, LSQ+ [1], was accepted to the EDLCV Workshop at CVPR 2020. After the competition, I continued to work on improving the performance of our unstructured pruning algorithm which led to Learned Threshold Pruning [2], a gradient-based method for learning per-layer pruning thresholds. In this work, although we achieved as high as 26× compression on AlexNet, this didn't translate to a 26× speedup on CPUs/GPUs because the unstructured nature of this sparsity can't be exploited by these platforms. This led to several considerations, such as using block-wise sparsity or using specific sparsity patterns that are amenable to hardware. Such hardware-software co-design has been a regular aspect of my research work at Qualcomm. Very recently, my work on Structured Kernels and Structural Regularization [3] was accepted as a full paper to NeurIPS 2020. Overall, my research has spread across several use-cases including image & video classification/segmentation, gaze estimation and hand-pose estimation. My near-term research goal is to extend the above ideas for designing low-complexity architectures capable of modeling long-range interactions across different components in 3D environments.

Apart from working on my ongoing research, I have also enjoyed working on several initiatives which have now developed into mature projects. While Qualcomm's participation in the MicroNet competition was first of my initiatives, I also started a small Computer Vision reading group which has now led to engagement from several teams in Qualcomm. Along with ensuring that the reading group sessions covered a diverse set of topics, I also made sure that these topics were aligned with the attending teams' business interests. This has given me a broader perspective of the problems that are currently relevant in the industry. As a result of my interactions with the AR/VR team, I initiated (and I currently lead) the Compute-Adaptive Perception track, where we aim to build models that can scale themselves, at runtime, with the changes in available resources. Concurrently, I am also mentoring/managing John Yang, summer intern, on the 3D hand-pose estimation problem. Working on these initiatives has taught me about creating effective roadmaps, articulating objectives and about being empathetic and creating a positive environment while doing the same.

Before Qualcomm, I got the opportunity to work with Prof. Jason Corso from University of Michigan in his start-up, Voxel51. I worked on developing real-time video detection+tracking algorithms to perform *querying* on large-scale video databases. Working on real-world data, I heavily engineered ideas for robust detection of blurred and/or tiny objects, frame-skipping, feature-reuse across frames, etc. to refine the predictions and

bring the processing speed close to real-time. This was truly an enriching experience as it brought a mix of both academia and industry - one one hand, I learned how to write bug-free production level pipelines from scratch and under tight customer deadlines; whereas I also worked with Prof. Corso on exploring several efficient backbones/techniques for object detection+tracking and activity classification on videos. Overall, I built two pipelines, *vehicle-sense* and *person-sense*, that are currently deployed on Voxel51's cloud service.

Apart from my work at Qualcomm and Voxel51, I have also pursued several projects and internships to upskill myself. My first deep learning related internship was at IBM Research - Bangalore, where I used a Common Representation Learning (CRL) based approach to build a fast catalog search engine for large fashion databases. At IIT Bombay, I pursued my Bachelors thesis under the guidance of Prof. Vikram Gadre in a collaboration with the Department of Cyber Security, Maharashtra, where I completed a comprehensive research study on using Scattering Wavelet Networks (ScatNets) for robust feature extraction and eventually used it for classification of latent fingerprints (i.e., raw imprints obtained from forensic documents). . I was awarded the **Undergraduate Research Award** (URA 02) by IIT Bombay for my thesis. One of my parallel work on ScatNets was during my internship with the Image and Signal Processing group at IFPEN, Paris on seismic sensor images - this work was presented at ICASSP 2018 [4].

While the above projects involved large-scale image & video datasets, I also pursued two projects that dealt with learning from limited or noisy data. The first was when I interned with the Watson Languages group at IBM Research - Almaden, where I developed pipelines to train sentiment classification models using noisy labels. I implemented several ideas from semi-supervised learning and otherwise (e.g. the Noise Adaption Layer, ICLR 2016) and eventually proposed a Teacher-Student learning method based on a curriculum learning to tackle this problem [5]. The second project involved segmentation of anatomical structures in chest radiographs, where we proposed an LP-based active learning framework to utilize weaker forms of annotations (bounding boxes and landmark points) to train a segmentation model in a *mixed-supervision setting* [6].

Ever since my undergrad experiences at IIT Bombay as a TA for *Quantum Mechanics & Applications* and *Wavelets* courses, I have been very passionate about teaching. I was humbled to serve as a Graduate Student Instructor (GSI) for 2 different courses at UMich - *Computational Data Science (CDS)* and *Logic for Computer Science*. This was an experience where the more I taught, the more I learned. For example, in the CDS course under Prof. Raj Nadakuditi, we created engaging lectures using Jupyter notebooks in Julia that gave the students hands-on experience with a variety of algorithms. I also helped setup challenging projects that taught me about the logistics and different evaluation methods to help the students best learn their concepts. I was selected as one of the top nominees for the **Towner Prize for Outstanding GSIs** at UMich.

References

[1] Yash Bhalgat, Jinwon Lee, Markus Nagel, Tijmen Blankevoort, and Nojun Kwak. Lsq+: Improving low-bit quantization through learnable offsets and better initialization. In *Proc. of CVPR Workshops*, 2020.

[2] Kambiz Azarian, Yash Bhalgat, Jinwon Lee, and Tijmen Blankevoort. Learned threshold pruning. *arXiv preprint arXiv:2003.00075,* ***currently under review at ICLR 2021***.

[3] Yash Bhalgat, Yizhe Zhang, Jamie Lin, and Fatih Porikli. Structured convolutions for efficient neural network design. In *Advances in Neural Information Processing Systems*, 2020.

[4] Y. Bhalgat, J. Charléty, and L. Duval. Catseyes: categorizing seismic structures with tessellated scattering wavelet networks. In *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2018.

[5] Yash Bhalgat, Zhe Liu, Pritam Gundecha, Jalal Mahmud, and Amita Misra. Teacher-student learning paradigm for tri-training: An efficient method for unlabeled data exploitation. In *KONVENS*, 2019.

[6] Yash Bhalgat, Meet Shah, and Suyash Awate. Annotation-cost minimization for medical image segmentation using suggestive mixed supervision fully convolutional networks. In *Medical Imaging meets NeurIPS workshop*, 2018.