Yash Sanjay Bhalgat

Senior ML Researcher, Qualcomm AI Research

Research Statement

Webpage: `https://yashbhalgat.github.io`

My interest in pursuing a PhD 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 efficient inference. Rapidly emerging applications in autonomous vehicles and AR/VR have led to tremendous progress in algorithms for 3D perception, yet these often fall short of the latency and power requirements. With my current experience at Qualcomm AI Research working on the intersection of hardware and ML, I have developed a deeper understanding of the challenges posed by inference on resource-constrained devices. In addition, fundamental differences arise when we move from the 2D setting of images to non-Euclidean manifolds such as 3D point clouds or meshes. Besides Neural Architecture Search or model compression based approaches for efficiency, I aim to explore these differences in developing efficient geometric deep learning algorithms.

Furthermore, perception algorithms often fail to generalize to unseen environments. Incorporating inductive biases can not only help models generalize to challenging environments but also make them more parameter-efficient. To this end, I would like to explore self-supervised representation learning algorithms that incorporate various inductive biases available in 3D geometries to learn meaningful representations for 3D data or more generally multi-modal data (e.g. point clouds + images + various other sensor data).

In my past 2 years at Qualcomm AI Research, I have been privileged to work with and learn from amazing researchers from deep learning, hardware systems as well as wireless communications. My research has been focused on algorithm and system design to develop efficient deep networks for computer vision use-cases. As my very first project, I led a team that won **3$^{\text{rd}}$ position** in the MicroNet competition at **NeurIPS '19**. The goal was to create the most efficient model that can achieve a 75% top-1 accuracy on ImageNet. In the 2-month span of this competition, I implemented several model quantization and pruning methods from literature and eventually proposed a novel and effective method to quantize deep neural networks to extremely low bit-widths. I presented this low-bit quantization work, LSQ+ [1], at the Efficient Deep Learning Workshop in **CVPR '20**. 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. This work has been submitted and under review at ICLR '21. In this work, I realized that a 26$\times$ model size reduction of an AlexNet model doesn't really translate to a 26$\times$ speedup on CPUs/GPUs, because unstructured sparsity can't be exploited by these hardware 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 at Qualcomm. Most recently, I published my work in **NeurIPS '20** wherein I introduced a novel type of convolution, called "Structured Convolutions" [3]. This idea led to structured deep networks that are 2$\times$ more efficient than state-of-the-art as it leverages the fact that additions are much cheaper than multiplications.

I am someone who is passionate about working collaboratively and have worked on several initiatives that foster collaboration. At Qualcomm, I founded the Computer Vision reading group to encourage collaborations among peers. This group started with 8 members and has now gained engagement from several teams across Qualcomm. Along with ensuring diversity in the reading group topics, I also made sure that there was business alignment. This has given me a broader perspective of the problems currently relevant in the industry. I have also started the Compute-Adaptive Perception track as a result of my interactions with the AR/VR team, where the goal is to build models that can adapt, at runtime, with changes in available resources. Leading this track educated me in creating effective project roadmaps, setting milestones and articulating objectives. I am also currently mentoring an intern on 3D hand-pose estimation, where we are working on an adaptive model architecture for this problem. This mentoring experience has really taught me about being empathetic and creating a supportive atmosphere. It has also pushed me to understand my own limitations and grow better in terms of technical as well as leadership abilities.

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 and 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. I also learned a lot by working with Prof. Corso on exploring several efficient backbones and techniques for object detection+tracking and activity classification on videos. This was truly an enriching experience with a mix of academia and industry that trained me to work in a challenging and fast-paced environment.

Prior to joining industry, I acquired a Masters in Computer Science and a Bachelors in Electrical Engineering from University of Michigan and IIT Bombay, respectively. During this period, I pursued several projects and internships at **IBM Research** - Bangalore, where I used a Common Representation Learning based approach to build a fast catalog search engine for large fashion databases. During my internship at **IFPEN, Paris**, I worked with the Image and Signal Processing group on using Scattering Wavelet Networks (ScatNets) for extracting sparse representations and categorization of seismic sensor images. This work [4] was presented at **ICASSP '18**. I also used ScatNets in my Bachelors thesis on robust classification of latent fingerprints. I was awarded the **Undergraduate Research Award** by IIT Bombay for my thesis work which was done under the guidance of Prof. Vikram Gadre in a collaboration with the Department of Cyber Security, Maharashtra. During my Masters at UMich, I collaborated with my undergraduate colleagues on the problem of segmentation of anatomical structures in chest radiographs. In this project, we dealt with the problem of training with limited data. We proposed an LP-based active learning framework to utilize weaker forms of annotations (bounding boxes and landmark points) in a mixed-supervision setting. This work was presented at Medical Imaging meets **NeurIPS '18** workshop [5].

Outside research, I have served as a TA at IIT Bombay for *Quantum Mechanics* and *Wavelets* courses, and as a Graduate Student Instructor (GSI) at UMich for *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 with Prof. Raj Nadakuditi, I created engaging lectures and challenging projects using Jupyter notebooks in Julia which taught me about the logistics and evaluation methods to help students best learn their concepts. I was 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*.

[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, 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.