Schmidt Science Fellowship Research Proposal by Tarek Allam Jr.

Efficient Learned Image Reconstruction and Compression Algorithms for Real-time Medical Image Analysis

---

Early cancer detection is vital for improving survival rates and for minimising pain and suffering to patients. Yet, opportunities for early detection are often missed either through lack of accessibility to equipment, affordability, or even logistical constraints of a health service. I will create a more accessible and affordable medical imaging device that is able to perform real-time enhanced analysis for better diagnostics, and ultimately improved rates of early cancer detection. Through a co-design of device and algorithm with clinicians, engineers and researchers, I will develop technology that can sit at the point-of-care and be available to patients across the world, encouraging early detection of curable diseases.

The non-ionizing and generally painless imaging modality of ultrasound offers a method for diagnostics that is well suited for early cancer detection being less expensive and bulky than other imaging techniques. While there have been steady improvements to ultrasound device design over the years, there is still a desire for higher-fidelity images at reduced latency as this enables better visualisation of tissues and potential pathologies in real-time. The comparably high data-rates involved with ultrasound presents several engineering challenges that can be solved with the methods proposed here, leveraging advances in machine learning specific hardware and efficient deep learning algorithms. By leveraging learned image reconstruction (LIR) techniques and latest model and image compression algorithms, ultrasound has the potential to offer affordable imaging at the edge for all.

Modern Graphical Processing Units (GPUs) and machine learning specific hardware such as Tensor Processing Units (TPUs) have made possible real-time processing of large imaging pipelines. Of late, there has also been much development of frameworks that help facilitate hardware-accelerated real-time inference [1] and allow for on-device model deployment into not only GPUs and TPUs, but also in mobile and low-powered embedded systems [2]. Work by [3] sought to leverage these frameworks in conjunction with GPUs to achieve real-time processing of classical image reconstruction algorithms. Further, [4] also identified the power of GPUs for real-time customised visualisations in ultrasound imaging. Both works showcased the potential for hardware-accelerated operations for improved image reconstruction in ultrasound, and the possibility of including deep learning for LIR by virtue of such frameworks. Deep learning has come to dominate in nearly all areas of the imaging domain, with improved accuracy at lower runtimes compared to iterative methods, without compromising robustness [5]. In the ultrasound field, extensions of model-based iterative image reconstruction to a learned approach is also seeing success [6].

However, a completely integrated implementation of LIR algorithms in embedded systems for real-time analysis seems yet to be fully realised. If modern LIR algorithms can be placed at the edge by way of recent model compression and image compression techniques, and by leveraging latest developments in machine learning specific hardware, improved real-time ultrasound image analysis can be achieved. This innovative endeavour to place deep learning at the heart of the imaging device in the mobile or low-powered setting should pave the way for more accessibility among patients worldwide to receive early diagnosis.

To marry these worlds of embedded systems and biomedical imaging, novel LIR algorithms will need to be explored that take into account the constraints imposed by the entire system end-to-end. With my vision for on-device deployment in mind, application of techniques that enhance the data acquisition end of the pipeline [7] combined with model compression on the other end will ensure data rates are made low enough for fast reconstruction [8].

With my expertise in developing efficient deep learning architectures, I foresee myself sitting at the apex of machine learning hardware research and deep learning for image reconstruction galvanizing communities to tackle this problem together. My complimentary skills in software development, knowledge in machine learning for signal processing, and experience leading collaborative projects, gives me a competitive edge for research into this specific challenge and will ensure I deliver successfully.

To test my hypothesis, I will first survey the current landscape of LIR techniques for their feasibility in on-device deployment and real-time operations. This stage will involve discussions with clinicians to learn about the biomedical aspects that are most important when evaluating the quality of a reconstructed image. I will learn of artefacts that may occur as a result of imaging different tissues within the human body.

Following a satisfactory overview of the different deep learning approaches, I will build upon my masters experience to commence development of novel learned imaging algorithms co-designed for the hardware it will operate on. This will require me to learn about the subtleties of running on embedded systems and the challenges involved with real-time processing. I will discuss with those in the embedded systems community, as well as learning from those who have built similar resource constrained devices before. Their insight will help me to acquire the skills necessary for on-device deployment of efficient models.

Through an open-source approach to the project, I will actively encourage collaborative working between disciplines and I expect to eventually deliver firmware capable of high-fidelity image reconstruction operating in real-time for early cancer detection. If model and image compression techniques can also be fully harnessed, I can see the possibility for devices to move towards a wireless setting which could allow for further processing to take place in the cloud, making ultrasound imaging more portable. In time, I envisage a device capable of cancer detection to sit at the desk of every dentist or doctor worldwide, and is the mission I will dedicate my career towards achieving.

Exploration of this new space combined with my personal motivations and skill-set, makes for an exciting adventure. The Schmidt Schmidt Fellowship gives me the opportunity to make the first steps towards becoming a leader in the area of improved smart diagnostics by enabling others to make advances thanks to the open source nature of the project, driving progress in the field.

[1] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, *et al.*, "Tensorflow: Large-scale machine learning on heterogeneous distributed systems," *arXiv preprint arXiv:1603.04467*, 2016.

[2] R. David, J. Duke, A. Jain, V. J. Reddi, N. Jeffries, J. Li, N. Kreeger, I. Nappier, M. Natraj, S. Regev, *et al.*, "Tensorflow lite micro: Embedded machine learning on tinyml systems," *arXiv preprint arXiv:2010.08678*, 2020.

[3] P. Jarosik, M. Byra, and M. Lewandowski, "Waveflow-towards integration of ultrasound processing with deep learning," in *2018 IEEE International Ultrasonics Symposium (IUS)*, pp. 1–3, IEEE, 2018.

[4] D. Hyun, Y. L. Li, I. Steinberg, M. Jakovljevic, T. Klap, and J. J. Dahl, "An open source gpu-based beamformer for real-time ultrasound imaging and applications," in *2019 IEEE International Ultrasonics Symposium (IUS)*, pp. 20–23, IEEE, 2019.

[5] M. Z. Darestani, A. Chaudhari, and R. Heckel, "Measuring robustness in deep learning based compressive sensing," *arXiv preprint arXiv:2102.06103*, 2021.

[6] D. Hyun, A. Wiacek, S. Goudarzi, S. Rothlübbers, A. Asif, K. Eickel, Y. C. Eldar, J. Huang, M. Mischi, H. Rivaz, *et al.*, "Deep learning for ultrasound image formation: Cubdl evaluation framework & open datasets," *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, 2021.

[7] O. Drori, A. Mamistvalov, O. Solomon, and Y. C. Eldar, "Compressed ultrasound imaging: from sub-nyquist rates to super-resolution," *arXiv preprint arXiv:2107.11263*, 2021.

[8] Y. Cheng, D. Wang, P. Zhou, and T. Zhang, "A survey of model compression and acceleration for deep neural networks," *arXiv preprint arXiv:1710.09282*, 2017.