Research Statement:
Increasing the Robustness and Utility of Deep Learning-based Medical Image Reconstruction

Alan Q. Wang, *Cornell University*

Advisor: Mert R. Sabuncu
Committee: Chris Xu, Jayadev Acharya

August 2021

Introduction and Background

Medical imaging is an invaluable tool in both clinical and research settings. Different imaging modalities capture different physiological and biological phenomena at different spatial and temporal scales. For example, microscopy and its variants (such as fluorescence, multi-photon, or structured illumination) can be used to analyze molecular events at the sub-cellular level. As another example, magnetic resonance imaging (MRI) and positron emission tomography (PET) scans can be used to detect cancerous tumors or subtle changes due to pathology, and give doctors important early information about progressive diseases like Alzheimer's.

While considerable advances have been made in improving the quality of the images measured in different modalities, the signal-to-noise ratio (SNR) of these measurements is invariably constrained by factors like inherent noise in the measurement process, inaccuracies or aberrations in the technology, or fundamental limits in the underlying physics. For example, fluorescence microscopy is constrained in resolution due to diffraction limits and scattering effects in tissue, requiring long exposure times at the risk of photo-damage. Similarly, mainstream magnetic field strengths used in MRI typically lead to low spin polarization, thus necessitating long acquisition times to achieve sufficient SNR. Both examples outline constraints specific to the modality that limit our ability to

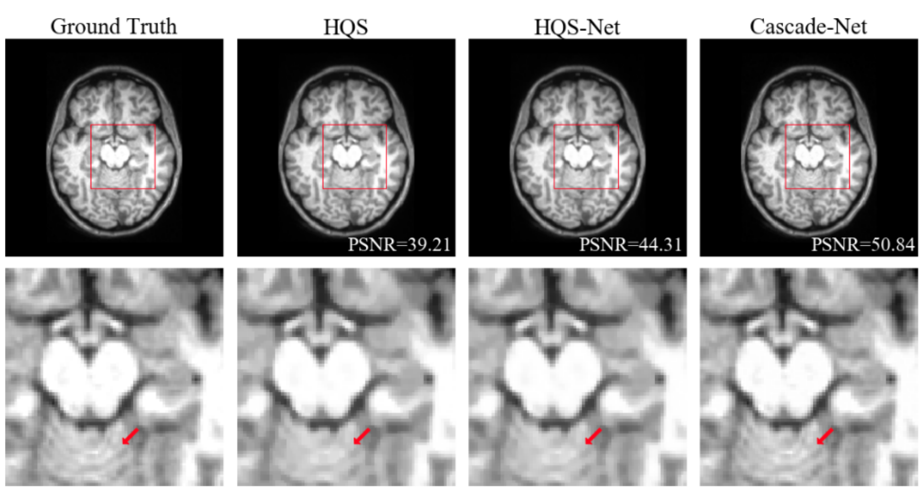

Figure 1: Representative brain slice of deep learning-based reconstruction for 4-fold under-sampled MRI scans. HQS denotes a non-deep learning based iterative algorithm, HQS-Net denotes our proposed unsupervised reconstruction model, and Cascade-Net denotes a supervised baseline model.

obtain high-quality images in a fast, cheap, and non-invasive manner. Thus, the task of recovering high quality images from noisy measurements, often called *image reconstruction*, is of crucial importance in medical imaging.

An increasingly-popular means of solving image reconstruction is via machine learning, and in particular deep learning. Broadly, this line of research seeks to use computational methods to improve the quality of the images, leveraging large datasets in order to train highly-parameterized and nonlinear neural network models. These reconstruction networks, which we will abbreviate as recon-nets, have shown impressive performance in a wide range of medical imaging applications. In regards to the two previous examples, recon-nets have been developed to perform image de-noising of microscopy images, effectively surpassing the diffraction limit in optical imaging [11]. Likewise, recon-nets have been used to remove artefacts from under-sampled MRI scans, opening up the possibility of decreased acquisition times with minimal sacrifice to imaging quality [5].

While recon-nets have demonstrated considerable improvement over more conventional reconstruction algorithms, gaps exist in the literature which hinder these models' mass deployment in clinical settings. One gap is the lack of research on recon-nets in *data-limited* scenarios. Presently, much of the success of deep learning is attributed to the supervised learning paradigm, where a large number of pairs of noisy and clean images are available for training. Such a dataset of necessary quantity and quality is often hard to obtain. Closely related to this is the lack of *model robustness* amongst supervised recon-nets. Research has shown that these variants perform poorly when the training and test data are not sampled from the same distribution (e.g. if noise is present at test-time), thus limiting model robustness and decreasing confidence in the efficacy of these models in the real world [1, 3, 6].

A second gap in the literature is the lack of *interpretable* recon-nets that put *humans in the loop*, which would allow human experts (e.g., radiologists) to oversee and intervene in model predictions [4, 10]. Such a paradigm is "the best of both worlds" in the sense that recon-nets can lessen the human burden and human experts can mitigate potential errors made by the recon-nets. As a motivating example, consider the results of Facebook's recent fastMRI Image Reconstruction Challenge[1], which revealed that commonly-used supervised metrics (e.g. mean-squared-error (MSE)) do not align with the quality of images as judged by radiologists. In a future where deep learning and human experts work in a mutually-beneficial way, the necessity for recon-nets which give nuanced and diverse predictions (rather than a single point prediction that simply achieves lowest MSE) is evident.

Past Research

In the first two years of my PhD program, I have made preliminary strides to address these gaps. As a way to address the first gap of data-limited training, we proposed and developed one of the first works tackling unsupervised reconstruction in compressed sensing MRI (CS-MRI), which seeks to develop models that reconstruct under-sampled MRI scans without the use of clean, fully-sampled data during training. Our first work to this end was HQS-Net, an unsupervised recon-net with an unrolled architecture, which performed amortized optimization of a classical regularized regression cost function over a dataset of under-sampled measurements. We were able to demonstrate superior robustness under noisy scenarios as compared to supervised baselines, as well as superior performance compared to classical non-deep learning-based techniques. An example of reconstruction performance is shown in Figure 1. This work was published in the International Workshop on Machine Learning for Medical Image Reconstruction 2020 [7].

We extended this work to address the second gap by injecting human-in-the-loop capabilities to the unsupervised reconstruction setting. We were initially motivated in this direction because of a perceived drawback of HQS-Net; namely, that the quality of reconstructions of the recon-net was heavily influenced by a hyperparameter which weighed competing terms in the loss function. To overcome this, we proposed and developed HyperRecon that uses a so-called hypernetwork to

[1] https://ai.facebook.com/blog/results-of-the-first-fastmri-image-reconstruction-challenge

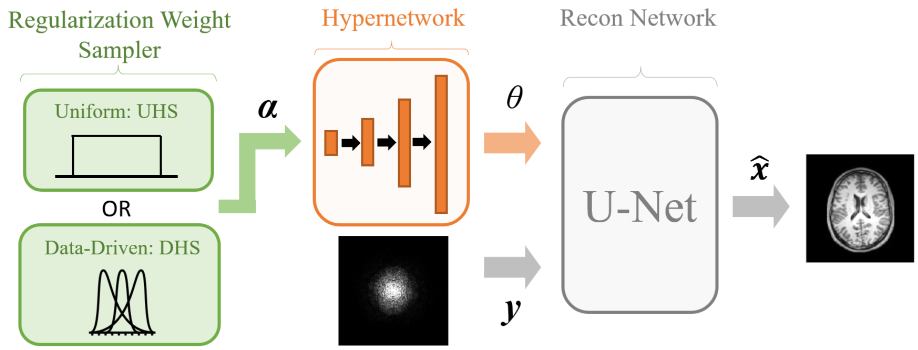

Figure 2: Proposed architecture of HyperRecon [8], which uses a hypernetwork to produce a rich set of diverse and plausible reconstructions at test-time.

take as input a specific hyperparameter value and output the weights of the recon-net, effectively making the hyperparameter an input to the model. This dual architecture, depicted in Figure 2, makes it computationally possible to compute a large number of reconstructions at test time – each corresponding to a specific hyperparameter value. These reconstructions can in turn be made available to a human user who can visually inspect them to choose the one/ones that they prefer. As such, this model allowed end users to effectively "turn a knob" and see in real time a rich set of diverse and plausible reconstructions. Indeed, we found that a number of reconstructions which had similar image-level metric performance (e.g. MSE) were quite different visually; in a standard recon-net model, only one of these optimal reconstructions would be available and the other equally-optimal reconstructions would be completely ignored. This work is currently under review, but a preprint is publicly available [8].

In a separate line of work, we sought to address the second gap of interpretability from a different direction; namely, what "role" do recon-nets serve in the imaging process? In the compressed sensing paradigm, the first phase of imaging involves collecting under-sampled measurements in an alternative basis (known as sensing), and the second phase involves reconstruction of these measurements to produce the final image. We were curious about whether all types of sensing information are equal, and whether or not recon-nets can do more with "better" sensing information. To explore this, we proposed and developed two works involving end-to-end optimization of sensing and reconstruction: the first in MRI and the second in fluorescence microscopy. While recon-nets had been well-explored in the literature at the time, we were the first to propose optimizing the sensing scheme jointly with the recon-net, thus enabling better overall performance than if they were to be optimized separately. In addition, the way the end-to-end model learned to delegate roles in both the sensing and reconstruction phases were of particular interest, as we hoped this would shine light on how to interpret the role of recon-nets. What information from the sensing phase would be most crucial for recon-nets to receive as input in order to maximize reconstruction performance?

Both works involved designing a stochastic sensing scheme which modeled the measurement space and the locations to collect, with a constraint on the total number of measurements collected. These measurements were subsequently passed to a recon-net which was trained to minimize MSE against clean images. In the first work, the measurement space was the Fourier (frequency) space, while in the second work the measurement space was the Hadamard (sequency) space. A diagram of the model in the second work is depicted in Figure 3. In both works, we were able to demonstrate superior image quality over baseline methods, while drastically decreasing acquisition times and potential invasiveness (e.g., due to photo-damage). Furthermore, we found that the optimized end-to-end scheme collected mostly low-frequency/low-sequency information during sensing, and subsequently "assigned" the task of imputing high-frequency/high-sequency information to neural

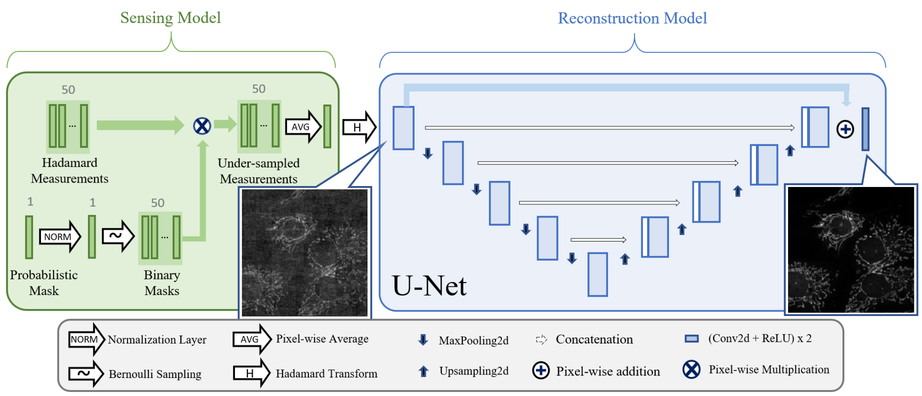

Figure 3: Proposed architecture of joint optimization of Hadamard sensing and reconstruction for fluorescence microscopy [9]. The stochastic sensing and U-Net based reconstruction model is optimized end-to-end.

networks during reconstruction, offering a novel insight into the capabilities of recon-nets. A journal article of the first work was published in IEEE Transactions on Computational Imaging [2], and a conference submission of the second work will be published at the International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI) 2021 [9].

Future Research

While these works have begun to address some of the aforementioned gaps in the literature, several future directions remain to be explored. Our approach to optimizing the sensing scheme in compressed sensing is a promising direction to increasing performance and interpretability in many other modalities. Although each modality has unique constraints and issues that need to be addressed, all would benefit from optimized sensing schemes in order to more intelligently collect measurements, which in turn would reduce costs and increase safety and efficiency. For example, in close collaboration with our colleagues at Weill Cornell Medicine, we have conducted preliminary experiments on under-sampled PET scans and have shown promise in both accelerating acquisition as well as improving image quality. Similar experiments could be applied to structural illumination microscopy or single-pixel imaging; i.e. which patterns of spatially-varying illumination result in the best imaging quality under a total-photon or acquisition-time constraint?

On another front, hypernetworks and the groundwork of HyperRecon are a potentially powerful means of addressing human-in-the-loop approaches to machine learning in medical imaging. While HyperRecon addressed the unsupervised setting, our preliminary results indicate that such a method works well in the supervised setup too, indicating the potential for a single model to provide high-quality reconstructions for any combination of supervised losses in real time. This approach will give us the flexibility to be agnostic to the choice of loss functions in training our models. Furthermore, at deployment, the end user will be able to quickly view a diverse set of reconstructions, enabling critical human oversight and intervention. In future work, we plan to address several open questions: How do we design architectures to maximize the expressiveness of hypernetworks? Once we have this rich set of predictions at our disposal, how do we choose the most salient and diverse reconstructions to present to the human user? Finally, are there ways of using hypernetworks to inject diversity and variability beyond loss functions?

In the remainder of my PhD, I plan to conduct more research exploring these directions. In further development of the first line of work, I plan to work closely with domain experts in different modalities to understand the imaging technologies in order to develop better models. In pursuit of the second line of work, I plan to further understand and optimize hypernetwork architectures for the reconstruction problem. I then plan to design methods of measuring salience and diversity of reconstructions so that they can be filtered into the most important reconstructions to show to end users.

References

[1] Vegard Antun, Francesco Renna, Clarice Poon, Ben Adcock, and Anders C. Hansen. On instabilities of deep learning in image reconstruction and the potential costs of ai. *Proceedings of the National Academy of Sciences*, 117(48):30088–30095, 2020.

[2] Cagla D. Bahadir*, Alan Q. Wang*, Adrian V. Dalca, and Mert R. Sabuncu. Deep-learning-based optimization of the under-sampling pattern in mri. *IEEE Transactions on Computational Imaging*, 6:1139–1152, 2020.

[3] Mohammad Zalbagi Darestani, Akshay S. Chaudhari, and Reinhard Heckel. Measuring robustness in deep learning based compressive sensing. *arXiv preprint*, 2021.

[4] Andreas Holzinger. Interactive machine learning for health informatics: when do we need the human-in-the-loop? *Brain Informatics*, 3(2):119–131, Jun 2016.

[5] Jin Liu, Yi Pan, Min Li, Ziyue Chen, Lu Tang, Chengqian Lu, and Jianxin Wang. Applications of deep learning to mri images: A survey. *Big Data Mining and Analytics*, 1(1):1–18, 2018.

[6] Xingjun Ma, Yuhao Niu, Lin Gu, Yisen Wang, Yitian Zhao, James Bailey, and Feng Lu. Understanding adversarial attacks on deep learning based medical image analysis systems. *Pattern Recognition*, 110:107332, 2021.

[7] Alan Q. Wang, Adrian V. Dalca, and Mert R. Sabuncu. Neural network-based reconstruction in compressed sensing mri without fully-sampled training data. In Farah Deeba, Patricia Johnson, Tobias Würfl, and Jong Chul Ye, editors, *Machine Learning for Medical Image Reconstruction*, pages 27–37, Cham, 2020. Springer International Publishing.

[8] Alan Q. Wang, Adrian V. Dalca, and Mert R. Sabuncu. Regularization-agnostic compressed sensing mri reconstruction with hypernetworks. *arXiv preprint*, 2021.

[9] Alan Q. Wang*, Aaron K. LaViolette*, Leo Moon, Chris Xu, and Mert R. Sabuncu. Joint optimization of hadamard sensing and reconstruction in compressed sensing fluorescence microscopy. *arXiv preprint*, 2021.

[10] Doris Xin, Litian Ma, Jialin Liu, Stephen Macke, Shuchen Song, and Aditya Parameswaran. Accelerating human-in-the-loop machine learning: Challenges and opportunities. In *Proceedings of the Second Workshop on Data Management for End-To-End Machine Learning*, DEEM'18, New York, NY, USA, 2018. Association for Computing Machinery.

[11] Fuyong Xing, Yuanpu Xie, Hai Su, Fujun Liu, and Lin Yang. Deep learning in microscopy image analysis: A survey. *IEEE Transactions on Neural Networks and Learning Systems*, 29(10):4550–4568, 2018.