Research Statement

Tao Hong From Large-Scale Computations to Reliable and Efficient AI: Advancing Imaging Science

I aim to develop provable and efficient computational methods for reliable AI-driven imaging science, enabling fast, high-quality imaging that increases patient throughput and accelerates scientific discovery.

Imaging is essential for advancing human health and scientific discovery by revealing the unseen. For example, biomedical imaging transforms invisible tissue and physiological processes into visual and quantitative information, enabling earlier disease detection, guiding life-saving treatments, reducing unnecessary procedures and costs, and improving long-term patient outcomes—thereby delivering broad benefits to healthcare and society. Biological imaging plays a complementary role at the cellular and molecular levels, revealing processes such as protein dynamics, neuronal activity, and cell–tissue interactions that underpin health and disease. High-dimensional imaging is essential because biomedical systems are volumetric and dynamic, while biological imaging reveals structural and molecular details from molecules to tissues. Due to factors such as hardware cost, radiation dose, patient comfort, and physical limitations, among others, the acquired data are often incomplete and noisy, making ***accurate***, ***robust***, and ***trustworthy*** reconstruction challenging. Moreover, *high-dimensional imaging* also poses challenges in ***computation*** and ***memory***. Fast computational methods are critical for time-sensitive care, such as emergency diagnosis and surgical guidance, and for reducing costs to enable use in resource-limited settings, thereby improving accessibility and impact.

My interdisciplinary background and research approach (see Fig. 1), integrating computational imaging, scientific computing, large-scale optimization, and trustworthy AI, places me in a unique position to advance scalable and reliable AI-driven methods for next-generation biomedical and biological imaging.

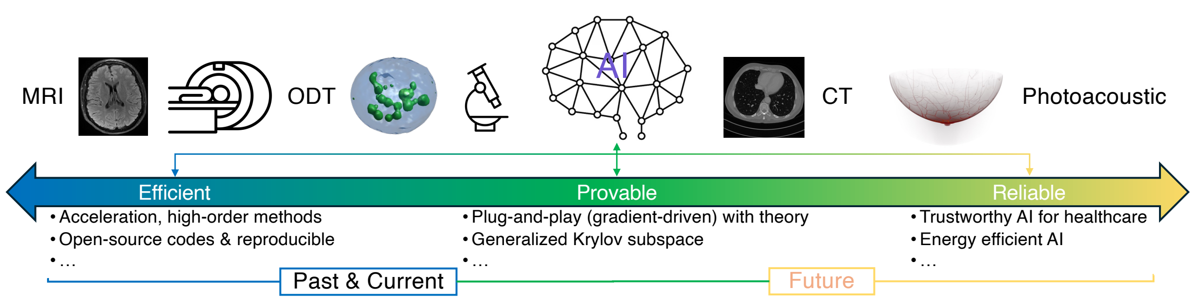

Figure 1: Research vision: large-scale optimization → provable AI → trustworthy, efficient AI.

1 Prior Research Experience

My training in large-scale optimization, scientific computing, AI, and theory equips me to tackle computational challenges in large-scale imaging and to develop efficient, reliable AI models for imaging science. Below, I highlight examples of my current and past PhD research and their real-world impact.

1.1 *Large-Scale Computations and Reliable AI Models in Imaging Science (Current)*

I have led several impactful projects focused on advancing image reconstruction quality and efficiency through the integration of AI-driven priors with provable computational methods. *A central challenge is building computational methods that are efficient, theoretically sound, and capable of employing reliable AI-driven priors for large-scale imaging under indirect, noisy, and undersampled measurements.* My past research tackles three problems central to this challenge. **P1.** How can we improve the convergence speed of image reconstruction in the plug-and-play (PnP) framework? **P2.** How can we accelerate the convergence rates of gradient-driven denoisers based reconstruction beyond first-order methods? **P3.** How can we adopt randomized algorithms to accelerate variational image reconstruction?

In my most recent work [1], I addressed **P1** by introducing preconditioning techniques for PnP compressed sensing (CS) MRI reconstruction. Modern MRI scanners use multi-coil arrays, and coil sensitivity maps vary between scans, which makes it difficult for standard end-to-end AI models to generalize; in contrast, PnP methods with learned denoisers naturally adapt to different scans. However, PnP methods converge slowly, limiting their applicability to large-scale problems. I developed effective preconditioners with rigorous convergence analysis, demonstrating that preconditioning can be systematically integrated into the

PnP framework, enabling both faster convergence and provable guarantees. *This work established the first provable preconditioned PnP scheme, reducing reconstruction time, improving image quality, and opening principled avenues to integrate preconditioning into AI-driven imaging and other large-scale inverse problems.*

My work [2] considered **P2** by incorporating curvature information. A key challenge in PnP is developing efficient, convergent algorithms whose assumptions are easily satisfied by trained convolutional neural network (CNN) denoisers. Gradient-driven denoisers (GDDs) overcome this limitation by requiring only Lipschitz continuity of the denoiser, while matching PnP performance and yielding a differentiable but nonconvex minimization problem. However, first-order methods converge slowly in practice, especially for ill-conditioned CS-MRI problems. Leveraging the fact that MRI images are complex-valued, I developed a *complex* second-order method by proposing a way to estimate a Hermitian positive-definite Hessian matrix in nonconvex settings. This approach attains much faster convergence than competing methods while retaining rigorous guarantees under only a Lipschitz-continuity assumption on GDDs. *This work introduced a second-order framework for AI-driven reconstruction with provable convergence and motivated my recent work [3] on large-scale problems with learned priors—a direction I will continue to pursue.*

My work on randomized algorithms [4] addressed **P3** by computing a preconditioner to accelerate variational image reconstruction using the Nyström approximation (NA). A key challenge in designing preconditioners for image reconstruction is twofold: the forward model is operator-based, and the preconditioner must be computed on-the-fly. I observed that the NA provides a promising preconditioner and demonstrated how it can be incorporated to accelerate existing algorithms for image reconstruction with impulsive or Gaussian noise. By leveraging modern GPU platforms, I further showed how to compute the preconditioner on-the-fly. The results demonstrated that my approach reduced reconstruction time from one hour to one minute.

1.2 *Efficient Methods in Imaging Science and Scientific Computing (PhD Research)*

My PhD research focused on developing efficient algorithms for large-scale problems, e.g., high-order methods for linear and nonlinear inverse problems [5, 8], multigrid-based solvers for diffraction tomography [9], hybrid approaches that combined multigrid optimization with sequential subspace optimization to balance global exploration with local refinement [10], and acceleration methods [6, 7] etc. My research addresses the following question: *How can we develop efficient, theoretically grounded algorithms to accelerate convergence, improve robustness, and solve large-scale inverse problems?*

My work in [5] addressed 3D optical diffraction tomography (ODT) reconstruction—a large-scale, nonlinear problem that is central to computational microscopy. A key challenge in 3D ODT arises from the nonlinear forward model required for high refractive index objects, which renders the problem highly nonconvex and both computationally and memory intensive. To mitigate the challenge, I developed a mini-batch quasi-Newton proximal algorithm that combines stochastic gradient updates with curvature information estimated from noisy gradients. This method converges faster than stochastic first-order approaches in both *iterations* and *wall time*, while preserving rigorous convergence guarantees. Validation on 3D *real* data further demonstrated its advantages. *This work pushed high-order methods beyond theory, demonstrating practical impact in real-world applications and laying the foundation for part of my current research.*

My NLAA paper [6] generalized Nesterov's scheme (NS) to accelerate iterative methods (IMs) for linear equations $\mathbf{A}\mathbf{x} = \mathbf{f}$, where $\mathbf{A}$ is a sparse, large-scale, and ill-conditioned matrix [14, 17]. Compared with classic Krylov subspace methods, NS requires more iterations but offers simpler implementation, lower memory usage, and similar per-iteration costs as the corresponding unaccelerated IMs. While NS has been applied to gradient descent [15], proximal gradient descent [16], and Newton methods [18], its generalization to IMs remains unresolved. I showed that NS can indeed accelerate IMs, derived a closed-form solution for the optimal momentum, and identified a class of IMs where NS cannot be applied. *This work demonstrated the feasibility of generalized acceleration methods for IMs, while advancing them to tackle nonlinear problems remains an open challenge that I plan to pursue further.*

My other PhD research reflects my problem-solving philosophy: *once you truly understand a problem, you can solve it*. For example, in optimizing sensing matrices for robust CS systems [11], I found that optimized sensing matrices performed worse than random sampling. I traced this issue to sparse representation errors (SREs), which naturally occur in practical signals. By accounting for SREs, I developed an effective strategy that outperformed random sampling and *opened new perspectives for optimizing MRI sampling trajectories.*

Future Research

With training in prestigious research groups, I developed an interdisciplinary background bridging theory, computation, AI, and imaging science. Ready to hit the ground running as an assistant professor, I aim to advance my research through the following unique **D**irections.

D1. Designing Provable AI-Driven Methods for Large-Scale Imaging Science. Biomedical imaging problems are often high-dimensional, involving 3D images, 4D dynamic images, and even 5D images that incorporate space, time, and spectral or parametric dimensions. AI-driven priors encode learned structure and can improve fidelity beyond classical methods, yet many applications lack high-quality training data, necessitating realistic simulations, which I plan to pursue in future work. Since these reconstructions inform clinical and scientific decisions, principled theory and guarantees are essential to ensure safety and reliability. My goal is to develop provable computational methods for AI-driven inverse problems, enabling efficient optimization and learning, as exemplified in [1–3, 5]. Leveraging my interdisciplinary background, I aim to address large-scale real-world imaging problems—such as MRI, photoacoustic computed tomography, and Fourier ptychography—by developing methods that enhance healthcare, accelerate scientific discovery, and, in parallel, contribute open-source software that empowers the research community.

D2. Understanding Uncertainty in AI-Driven Methods to Enable Reliable Deployment. Clinicians rely on reconstructions for disease diagnosis. Therefore, we not only need guarantees on the reconstruction steps, but also must evaluate whether AI-driven priors can be trusted. Studying uncertainty in AI-driven methods is essential for reliable deployment. I plan to study uncertainty in biomedical applications (e.g., reconstruction and medical foundation models) using conformal inference methods (such as conformal prediction and conformal risk control), while also exploring complementary approaches to robust and trustworthy AI. Beyond supporting the development of reliable and safe AI models, a deeper understanding of uncertainty can shed light on fundamental challenges in AI-driven methods, including distribution shifts, bias, data corruption and adversarial attacks, hallucinations, and robustness—directions I also intend to explore.

D3. Enhancing Energy Efficiency of AI-Driven Methods. Given the large size of modern deep neural networks (DNNs), efficiency is essential for faster performance, reduced energy consumption, and sustainable (green) AI. This need is particularly critical for deployment on portable devices with limited power, memory, and computation. Parameter quantization provides a promising strategy for compressing networks and reducing computation [19]. Since matrix–vector multiplication (MVM)—a fundamental DNN operation—admits infinitely many realizations that affect quantization error (QE) [20], I plan to investigate realizations that minimize QE, building on my prior work with state-space methods in digital filters [12, 13] and earlier hardware research experience. Notably, state-space methods have also shown promise in time-sequence AI models [21, 22]. More broadly, my goal is to establish theoretical frameworks that guide efficient realizations across neural networks and to collaborate with hardware experts to enable deployment on edge devices, smartphones, and embedded systems. This direction forms a key pillar of my broader vision: developing efficient, robust, and trustworthy AI-driven methods for large-scale inverse problems.

Conclusion: I plan to lead a research program that not only advances the theory of computational methods and AI but also delivers real-world impact—improving imaging to advance healthcare and accelerating scientific discovery, while enabling trustworthy AI across science and engineering.

References

Part I – My Selected Publications (__ indicates corresponding author)

[1] **Tao Hong**, Xiaojian Xu, Jason Hu, and Jeffrey A. Fessler, Provable preconditioned plug-and-play approach for compressed sensing MRI Reconstruction, *IEEE Transactions on Computational Imaging*, vol. 10, pp. 1476 - 1488, Oct. 2024. [link] [pdf] [code]

[2] **Tao Hong**, Zhaoyi Xu, Se Young Chun, Luis Hernandez-Garcia, and Jeffrey A. Fessler, Convergent complex quasi-Newton proximal methods for gradient-driven denoisers in compressed sensing MRI reconstruction," *To appear in IEEE Transactions on Computational Imaging after Major Revision*, 2025. [link] [pdf] [website] [code]

[3] **Tao Hong**, Umberto Villa, and Jeffrey A. Fessler, A convergent generalized Krylov subspace method for compressed sensing MRI reconstruction with gradient-driven denoisers," *submitted to IEEE Transactions on Computational Imaging*, 2025. [link] [pdf]

[4] **Tao Hong**, Zhaoyi Xu, Jason Hu, and Jeffrey A. Fessler, On adapting randomized Nyström preconditioners to accelerate variational image reconstruction," *To appear in IEEE Transactions on Computational Imaging after Major Revision*, 2025. [link] [pdf] [code]

[5] **Tao Hong**, Thanh-an Pham, Irad Yavneh, and Michael Unser. A mini-batch quasi-Newton proximal method for constrained total-variation nonlinear image reconstruction. *Submitted to SIAM Journal on Imaging Sciences,* 2025. [link] [pdf] [poster] [code]

[6] **Tao Hong** and Irad Yavneh. On adapting Nesterov's scheme to accelerate iterative methods for linear problems. *Numerical Linear Algebra with Applications*, 29(2):e2417, 2022. [link] [pdf] [slides] [code]

[7] **Tao Hong**, Yaniv Romano, and Michael Elad. Acceleration of RED via vector extrapolation. *Journal of Visual Communication and Image Representation*, 63:102575, 2019. [link] [pdf] [code]

[8] **Tao Hong**, Irad Yavneh, and Michael Zibulevsky. Solving RED with weighted proximal methods. *IEEE Signal Processing Letters*, 27:501-505, 2020. [link] [pdf] [slides] [code]

[9] **Tao Hong**, Thanh-an Pham, Eran Treister, and Michael Unser. Diffraction tomography with Helmholtz equation: Efficient and robust multigrid- based solver. *In preprint*. [link] [pdf]

[10] **Tao Hong**, Irad Yavneh, and Michael Zibulevsky. Merging multigrid optimization with SESOP, *In preprint*. [link] [pdf] [slides]

[11] **Tao Hong** and Zhihui Zhu. An efficient method for robust projection matrix design. *Signal Processing,* 143:200-210, 2018. [link] [pdf] [code]

[12] **Tao Hong**, Chaogeng Huang, Gang Li, and Yong Ching Lim. A Hessenberg-based input balanced realization for all-pass systems. *9th International Conference on Information, Communications & Signal Processing.* IEEE, 2013. [pdf]

[13] **Tao Hong**, Si Tang, Gang Li, Xiongxiong He, and Liping Chang. All-pass based efficient and robust structures for finite precision implementation of digital filters. *Proceedings of the 32nd Chinese Control Conference,* IEEE, 2013. [pdf]

Part II – Other Publications

[14] Yousef Saad. *Iterative Methods for Sparse Linear Systems.* SIAM, 2003.

[15] Yurii Nesterov. A method for solving the convex programming problem with convergence rate $O(1/k^2)$. In *Dokl. akad. nauk Sssr,* 269:543-547, 1983.

[16] Amir Beck and Marc Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM Journal on Imaging Sciences,* 2(1):183-202, 2009.

[17] Achi Brandt. Multi-level adaptive solutions to boundary-value problems. *Mathematics of Computation*, 31(138):333-390, 1977.

[18] Yurii Nesterov. Accelerating the cubic regularization of Newton's method on convex problems. *Mathematical Programming,* 112:159–181, 2008.

[19] Amir Gholami, Sehoon Kim, and Zhen Dong et al. A survey of quantization methods for efficient neural network inference. *In Low-Power Computer Vision,* 291-326, 2022.

[20] Michel Gevers and Gang Li, *Parametrizations in Control, Estimation and Filtering Problems: Accuracy Aspects*. Springer Science & Business Media, 1993.

[21] Albert Gu and Tri Dao, Mamba: Linear-time sequence modeling with selective state spaces, *arXiv preprint arXiv:2312.00752, 2023.*

[22] Albert Gu, Isys Johnson, and Karan Goel et al. Mamba: Linear-time sequence modeling with selective state spaces, *Advances in Neural Information Processing Systems, 2021.*