SCIENCE CHINA
Information Sciences

• PERSPECTIVE •

November 2024, Vol. 67, Iss. 11, 216101:1–216101:3
https://doi.org/10.1007/s11432-024-4121-0

Embodied computational imaging: a new paradigm for observing and analyzing spatiotemporally ultrasensitive phenomena at multiple scales

Baoquan CHEN[1*], Zhouchen LIN[1], Peng XI[2], Yebin LIU[3] & Xiaodian CHEN[4]

[1]*School of Intelligence Science and Technology, Peking University, Beijing* 100871*, China;*
[2]*College of Future Technology, Peking University, Beijing* 100871*, China;*
[3]*Department of Automation, Tsinghua University, Beijing* 100084*, China;*
[4]*National Astronomical Observatories, Chinese Academy of Sciences, Beijing* 100101*, China*

Received 10 April 2024/Revised 11 June 2024/Accepted 5 August 2024/Published online 14 October 2024

Imaging is the foundational tool for scientists to observe and understand the world. It has served as the basis for scientific discoveries over human history. In recent years, with the rapid advancement of artificial intelligence and computing technologies, a new interdisciplinary field, computational imaging, has emerged and developed. By jointly optimizing imaging hardware and processing algorithms, computational imaging continuously breaks through the physical limits of traditional imaging systems in aspects such as resolution, dimension, contrast, and throughput. For instance, leveraging radio interferometry and compressed sensing reconstruction, the Event Horizon Telescope (EHT) Collaboration program successfully achieved a significant milestone by capturing the first photo of a black hole [1], providing important imaging support for Einstein's theory of general relativity. Adopting 3D reconstruction to the imaging process, the single-particle cryo-electron microscopy (cryo-EM) technique [2] substantially improves the precision of biological macromolecule structural analysis and provides a powerful tool for biology and drug design. Integrating advanced neural representations with transient imaging techniques facilitates nanosecond time-scale reconstruction, which allows for the recording of the photo propagation process, capturing the journey of light with unprecedented speed and accuracy. These computational imaging technologies bring new capabilities and significant potentials to scientific discoveries.

However, current imaging technology still faces severe challenges in observing phenomena with weak, transient signals that occur in vast spaces but are localized, emerge and develop rapidly, or vanish in a flash. We refer to this kind of phenomena as spatiotemporally ultrasensitive phenomena (SUP). Numerous scientifically significant phenomena fall into this category, including the operation of organelles and the transmission of biological signals within cells, supernovae and fast radio bursts in the cosmos, transient movements of animals or humans caused by neural activities, and rapidly evolving physical phenomena such as turbulence and explosions. These SUP, occurring across multiple scales, necessitate observation methods with ultra-high temporal and spatial resolution, the ability of continuous and real-time tracking, and precision in capturing rich dynamic details.

Existing imaging and computing systems are incapable of fulfilling these requirements. The major challenges are three folds:

- Slow to respond. Current imaging technologies typically operate passively to the changes of dynamic targets, relying heavily on human expertise in their deployment and parameter control. Limited by the cost and capabilities of sensing equipment, these imaging systems often struggle to keep up with SUP by timely deploying the rightful imaging devices and tuning the most suitable parameters. This results in low-quality imaging outcomes, significantly hampering scientific discovery. Therefore, there is an urgent need to design an end-to-end imaging system capable of predicting the emergence and evolution of SUP and proactively adjusting the imaging systems for capturing the phenomena with the highest fidelity.
- Hard to register. To capture SUP, it typically involves heterogeneous and multi-source signals, governed by various physical principles and processes. The resulting imaging data are multimodal in nature and scattered across multiple spatial/temporal regions/ranges. Therefore, fusing (registering) all these signals to form (reconstruct) a complete picture of the phenomena is extremely difficult. Therefore, there is an urgent need for data fusion, for which a coherent and precise physical modeling of the target phenomena is needed. This model needs to be differentiable to facilitate the proactive imaging approach, aided by efficient event prediction, capturing, and data fusion.
- Difficult to comprehend. SUP are typically governed by multiple, yet interconnected mechanisms with complex causal relationships. Existing machine learning methods lack the generality required to adapt to various SUP, to dissect their underlying principles, to further discover new sciences. Therefore, there is an urgent need for learning algorithms that can cater to dynamic phenomena. This means developing new learning algorithms that are online, continuous, and all the while, differentiable, is, again, necessary for the proactive and close-loop imaging approaches.

* Corresponding author (email: baoquan@pku.edu.cn)

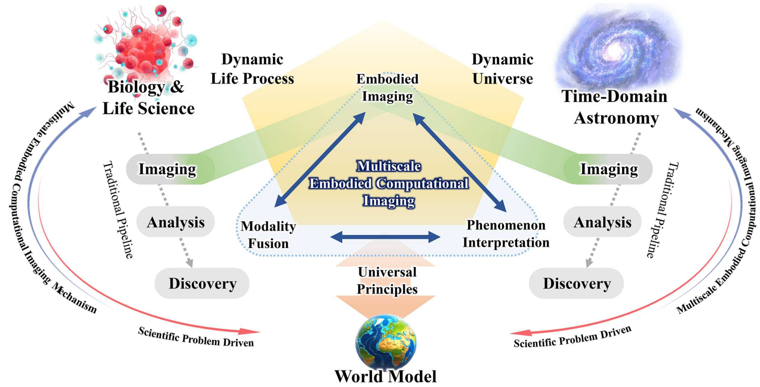

Figure 1 (Color online) Embodied computational imaging paradigm: traditional computational imaging pipeline takes imaging, analysis, and scientific discovery as sequentially separate stages, hindering the potential for seamless collaboration among them. The new paradigm, embodied computational imaging, treats the process from imaging to scientific discovery as an integrated, closed-loop pipeline. This new approach brings the opportunity of handling SUP, leading to revolutionary applications across various fields, such as biology, life sciences, and time-domain astronomy.

The scientific community has been shifting its focus from reductionism (taking the pieces apart) to systems theory (understanding the larger picture by putting its pieces together). For instance, the biology paradigm has evolved from molecular biology to integrative biology, and astronomy study has entered a new era of time-domain astronomy survey. This shift demands significant steps toward in situ imaging, holistic data integration, and adaptive learning frameworks. These are necessary to comprehensively capture and interlink, and unveil the multiple facets of physical information inherent in scientific phenomena and the fundamental laws governing them. However, within the current paradigm of computational imaging, imaging and scientific discovery are treated as two separate stages. The connection between them heavily relies on manual analysis and induction. This artificial separation introduces a barrier between event capturing and scientific discovery. Consequently, this affects the quality of imaging, the efficacy of analysis, and the capability of discovery.

Embodied computational imaging. To overcome the technological bottlenecks in multi-domain, multiscale scientific observation and discovery, we propose a new paradigm, embodied computational imaging, which addresses the common challenges mentioned above. Figure 1 illustrates the core concepts of this new paradigm.

Embodied computational imaging draws inspiration from embodied artificial intelligence (embodied AI), an emerging research field in artificial intelligence that has received widespread attention in recent years [3]. Embodied AI focuses on developing an understanding of the environment through embodied entities that engage in active interactions, thereby achieving the co-evolution of environmental understanding and interaction strategies. Similarly, in embodied computational imaging, imaging systems are no longer tools that respond passively but are intelligent agents equipped with built-in modeling and analysis abilities. These imaging agents are capable of sensing, computing, and taking actions. They autonomously optimize imaging strategies based on scientific problems and domain knowledge.

The core of embodied computational imaging is the trinity of embodied imaging, modality fusion, and phenomenon interpretation. The embodied imaging module upgrades the traditional imaging systems into differentiable embodied imaging agents, which autonomously intervene in the observation process, accept feedback from the other modules, optimize hardware design and parameters, and intelligently and proactively schedule observation. Synchronously, the modality fusion module models the evolution of SUP by fusing diverse, multimodal signals from multiple sources. It employs differentiable physical solvers to achieve efficient integration of physics priors and the registration of multimodal information. Meanwhile, built upon differentiable continuous active learning approaches and customizable machine learning frameworks, the phenomenon interpretation module leverages observation results and domain knowledge to analyze and interpret the physical laws underlying the phenomena, thereby acquiring new domain knowledge.

These three modules constitute a fully differentiable imaging and computing pipeline. Being differentiable implies that each component within the pipeline can produce predictions through embedded parametric models and can mutually fine-tune based on the discrepancies between their predictions and actual observations. Unlike existing methods that typically rely on pre-trained networks and cannot adaptively respond to new SUP, the proposed system features a tightly coupled software and hardware design, fitting in the framework of software-defined hardware. Each module of the system can optimize its parameters based on the outcomes of other modules. For example, in the system proposed by [4], the imaging process optimizes the hardware design and parameters on the fly based on observed phenomena and targeted scientific discovery, providing proactive and closed-loop sensing and interpretation, thus an overall significant performance improvement.

Embodied computational imaging effectively addresses the aforementioned challenges encountered by existing imaging systems when dealing with SUP. The Embodied Imaging module equips these systems with target-driven observation capabilities, enhancing their responsiveness to fast-changing environments and signals. The Modality Fusion module provides efficient ways to model physics and enables the registration of multimodal physical information to create a comprehensive view of complex phenomena. The Phenomenon Interpretation module significantly improves the adaptivity of the learning methods, enabling online understanding of the physical mechanisms behind phenomena. This entire paradigm operates in an integrated manner, enabling precise guidance in image acquisition and in situ observation, facilitating immediate analysis of observations, capturing the full scope of dynamic processes, and obtaining a comprehensive interpretation of SUP.

Prospects for life sciences and astronomy. The proposed embodied computational imaging paradigm enhances observational performance on SUP by supporting high fidelity imaging and effective data analysis, and fosters scientific discoveries across a broad spectrum of disciplines. It has the potential to revolutionize applications across various fields.

In biology and life science, the new paradigm can address the limitations of traditional optical microscopes and enable the creation of a next-generation microscopic imaging system capable of high-throughput, long-duration, high-resolution, multimodal, and multiscale imaging. Such a system can comprehensively capture morphological changes and interactions of living organisms across spatial and temporal scales. It potentially enables the creation of digital cells that precisely model the dynamic life processes of cells both spatially and temporally. This helps tackle significant problems in biology and life science research, including the mechanisms of signal storage and transmission in the brain, regulation, and counter-regulation in tumor cells, and comprehensive modeling of cellular processes.

Similarly, in time-domain astronomy, large-scale astronomical surveys play a critical role in the study of the dynamic universe, one of the three major scientific challenges in astronomy for the next decade. The new paradigm can transform telescope arrays into embodied imaging systems, allowing for intelligent and fully automatic observations that facilitate the systematic and rapid search for rare transient phenomena, thereby solving the challenges of high-frequency and large-scale surveys using large telescope arrays. Furthermore, the paradigm incorporates advanced methods for the detection and modeling of SUP. By integrating astronomical data across multiple wavelengths and messengers, it facilitates rapid and efficient follow-up precision measurements using more powerful telescopes. These complementary observations in time-domain and multi-messenger astronomy allow for the effective detection of previously elusive and rare transient phenomena. This improved observational capability opens new avenues for more detailed investigations into dark matter, dark energy, cosmic expansion, and physics in extreme environments such as black holes and stellar explosions.

Embodied computational imaging possesses a complete spectrum of capabilities, extending from the observation of phenomena to scientific discovery. This capability enables the continuous modeling of universal principles across various fields. With recent advancements in artificial intelligence, these universal principles can be encapsulated within what are referred to as World Models [5]. World Model is an emerging concept in both embodied AI and generative AI. When provided with a large amount of video footage and data from other modalities such as text, audio, and 3D information, a World Model is supposed to extract the physics and other principles embedded in the data, enable visually consistent content generation, and can serve as a key component in embodied AI systems that predict event evolution of the environment. Therefore, how to generate an accurate world model is of key importance.

With the capability of embodied computational imaging, the World Models can be continually built, evaluated, and updated with proactive live observations. The world models of various domains can be separately built, and together form a universal world model. For instance, by studying the motion and interaction across a continuum from the microscopic realm of molecules to the macroscopic levels of cells, tissues, organs, and entire living organisms, we can develop digital life process (DLP) models. These models offer a holistic view of the various stages and aspects of life, covering an impressive range of eight orders-of-magnitude in both time and space, and give rise to self-assembly cascades and intricate networks of interactions essential to the functioning of life. Similarly, by continuously observing and analyzing dynamic phenomena in the universe that occur over periods ranging from seconds to years, such as stars, white dwarfs, collisions of black holes, and stellar explosions, we can construct digital universe (DU) models. These models will revolutionize our comprehension of a wide range of subjects, from the origins of carbon in living cells and metals in our computers, to the universe's expansion history since the Big Bang, the lifecycle of stars, and the physics surrounding black hole event horizons.

Embodied computational imaging systems, equipped with proactive observation capabilities, are expected to offer an online framework for training and validating these domain-specific World Models. Moreover, World Models can guide embodied observation systems to enhance their capabilities via embodied exploration and self-improvement. By leveraging the accumulated domain knowledge and the World Models, embodied computational imaging systems will significantly increase the efficiency of observations and lead to significant scientific breakthroughs.

Conclusion. In this article, we have analyzed the common challenges that current imaging systems face, particularly when dealing with spatiotemporally ultrasensitive phenomena (SUP). We present embodied computational imaging, a new paradigm that addresses these challenges and has the potential to revolutionize applications across various fields.

Currently, the rapid development of artificial intelligence, such as embodied AI, differentiable physics, and customizable machine learning frameworks, has laid a solid foundation for realizing the concepts of embodied computational imaging. Meanwhile, the emergence of new theories, algorithms, and systems in artificial intelligence and world models enables high-quality imaging driven by large-scale data and rapid analysis, and furthermore, provides powerful tools for efficient learning of domain knowledge. Concurring with all of these advancements, developing the theory and system of embodied computational imaging will not only propel scientific discovery in various domains, but also help advance artificial intelligence in real world scenarios.

References

1 Akiyama K, Alberdi A, Alef W, et al. First M87 event horizon telescope results. I. The shadow of the supermassive black hole. Astrophys J Lett, 2019, 875: L1

2 Nogales E. The development of cryo-EM into a mainstream structural biology technique. Nat Methods, 2016, 13: 24–27

3 Driess D, Xia F, Sajjadi M, et al. PaLM-E: an embodied multimodal language model. In: Proceedings of the 40th International Conference on Machine Learning (ICML'23), Honolulu, 2023. 8469–8488

4 Chen W, Mirdehghan P, Fidler S, et al. Auto-tuning structured light by optical stochastic gradient descent. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2020), 2020. 5970–5980

5 LeCun Y. A path towards autonomous machine intelligence. OpenReview, 2022. https://openreview.net/pdf?id=BZ5a1r-kVsf