Research Statement

Jiayun Wang
http://pwang.pw

I develop machine learning (ML) and computer vision (CV) methods to address data insufficiency and model interpretability challenges. My techniques leverage physics simulations, reduce human supervision by learning from imbalanced data, and enhance interpretability through multi-level structure-aware representations. My goal is to incorporate domain knowledge into data-driven and interpretable models that not only perform well but also facilitate scientific discovery and provide insights for scientists and stakeholders.

Machine learning (ML) and computer vision (CV), specific forms of artificial intelligence, are gradually transforming various fields thanks to their applicability to numerous real-world problems. Notably, ML and CV have made tremendous developments in healthcare and is getting increasingly popular. By 2023, there are more than 23,000 articles on AI for healthcare on PubMed [1]. Nonetheless, despite the rapid increase in research papers, the deployment of MLCV models in real-world settings such as clinics, remains slow, particularly for high-stakes tasks. Several challenges impede the real-world application of MLCV models: **1) Limited and Imperfect Data:** Certain scenarios, like under-studied diseases or technological constraints of biomedical devices, suffer from limited data availability. Annotating medical data is costly. Since training AI models relies on high-quality data [2], the lack of such data can lead to model failures. **2) Model Interpretability:** Many MLCV models operate as black boxes, making it difficult for practitioners to assess the trustworthiness of their predictions.

I develop ML and CV methods that mitigate data/annotation insufficiency and low model interpretability issues. **1) Learning from Physics Simulations:** In cases where real data is unavailable, we leverage physics models and simulated data to develop ML. Our model trained on simulated data can effectively work on real organs. We proposed the **first method that reconstructs lung histological structure with commercial ultrasound device**. We also developed resolution-invariant methods that enable real-time photoacoustic imaging of the breast. **2) Reducing Human Supervision:** In cases where large-scale real data is available, significant human effort is needed to curate and annotate the data for training MLCV models. We develop methods to reduce human supervision by learning from imbalanced data and training models without annotations. Our approach has demonstrated superior performance with minimal human supervision in wildlife recognition and ocular disease diagnosis. **3) Enhancing Model Interpretability:** We improve model interpretability by learning multi-level structure-aware representations. Our models can explain the reasoning behind diagnoses of ophthalmic diseases and predictions of surgeon behaviors.

Through close collaboration with scientists and practitioners, my research aims to deeply integrate ML model development with real-world applications. By incorporating domain knowledge and data from experts, I create interpretable models that provide insights to facilitate scientific discovery.

Learning from Physics Simulation. Physics simulations are crucial in solving many complex scientific problems, especially when real data is limited but physical laws are well-understood. By modeling and simulation, researchers can understand intricate processes better, predict outcomes, and design more effective interventions across various scientific domains Medical imaging, one notable application, solves the inverse problem and creates visual representations of the interior of the body from observations of the medical device. To train MLCV models for medical imaging, a large-scale dataset of high-quality ground-

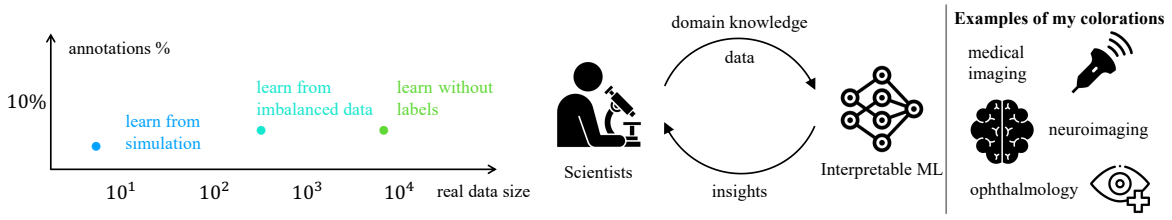

Figure 1: I develop interpretable ML and CV methods that learn from physics simulation, imbalanced and label-free data. In collaboration with scientists, our models show robust performance and extract insights for many applications.

truth images is necessary (e.g. knee reconstruction in MRI [3]). However, acquiring such images due to the intricate nature of biological systems and the limitations of current imaging technologies. For example, lung ultrasound imaging (LUS) is challenging due to the complexity of the wave propagation physics in the lung which exhibits complicated tissue-air structures. It is not possible to image lung histology in vivo, which developing data-driven models intractable.

We are changing the landscape of LUS and diagnosis by proposing the first method that reconstructs lung histology from ultrasound sensory data [4]. Since it is infeasible to obtain lung histopathology maps using in-vivo ultrasound, we collaborate with biophysicists to simulate ultrasound propagation in the lung. By leveraging these simulations, we develop inverse imaging techniques that, although trained on simulated data, demonstrate robust performance on real lungs.

Unlike other MLCV models which only operate on a specific resolution of data, our model is **resolution-invariant** and maintains consistent performance across varying resolutions of input measurements and output reconstructions. A direct application is compressed sensing, where our model can reconstruct high-quality images from sub-sampled measurements. Specifically in 3D photoacoustic imaging, we demonstrate the capability to reduce input measurement size and speed up the measurement time by up to 10x while maintaining minimal loss in reconstruction performance [5]. Our model enables **real-time 3D breast imaging** by enhancing temporal resolution. We also achieve high temporal-resolution functional ultrasound for brain imaging [6] with the same model design.

Learning from Imbalanced Data. A more suitable case for MLCV model development is when abundant data is available. However, this abundance of real-world data often comes with the challenge of imbalanced or long-tailed distributions. Imbalanced data would increase the bias towards the majority, hurt the performance of the minority (e.g. not detecting rare diseases), and even make the model collapse [2]. Rather than using resource-consuming data curation which needs heavy human supervision, we reduce supervision by proposing methods that achieve balanced performance for both the majority and minority.

We for the first time formally studied *Open Long-Tailed Recognition* for visual data [7], [8]. We formally **define the long-tailed recognition task, develop benchmark datasets and evaluation protocols, as well as a novel representation learning method** that outperforms state-of-the-art approaches. Our insight is that rather than learning from imbalanced labels, it is more appropriate to learn structure-aware representations from the data, which are regularized, shared across the dataset and not suffering the imbalanced issue. We apply the proposed framework to real-world applications including wildlife recognition [9] and ocular disease diagnosis [10] with balanced performance for both common and less-common classes.

Learning without Annotations. The abundance of real-world data also comes with the need for extensive human supervision and efforts for annotation. We adopt two methods to address the challenge: **1)** We directly learn from unlabeled data, which leverages unsupervised learning to extract useful features without manual labeling. We pioneer to propose the **first method that discovers joint object identity and pose** from unlabeled images [11]. **2)** Learning to distill and transfer knowledge from existing models also reduces the need of extensive annotations. Our method enables humans to collaboratively develop CVML models and shows strong applicability to predicting ophthalmic diseases [12] and surgery feedback [13].

Improving Interpretability. Interpretable ML models are pivotal in many medical and scientific applications for improving model transparency and trustworthiness. We leverage multi-level structure-aware representations for the morphology of meibomian glands [10], [14], use them to interpret the disease diagnosis ML model and identify the most important indicators for diagnoses [15]–[17].

Future Directions. My current research focuses primarily on single objects or localized views, but expanding to more complex scenarios involving more complicated systems is of significant interest. Future work will aim to develop geometry-aware representations for multi-object scenes, enabling more comprehensive and accurate modeling of intricate biological structures. Additionally, integrating multi-modality data, such as video and sound, will enrich the data landscape, providing a more holistic view of the studied phenomena [13]. Improving the efficiency of ML models to facilitate their practical deployment is also a key objective. Notably, we have developed lean and efficient MLCV models by preserving data structures such as orthogonality and recurrence [18], [19]. These advancements will enhance the applicability and performance of ML and CV technologies in real-world healthcare settings.

References

[1] R. Awasthi, S. Mishra, R. Grasfield, J. Maslinski, D. Mahapatra, *et al.*, "Artificial intelligence in healthcare: 2023 year in review," *medRxiv*, pp. 2024–02, 2024.

[2] I. Shumailov, Z. Shumaylov, Y. Zhao, N. Papernot, R. Anderson, *et al.*, "Ai models collapse when trained on recursively generated data," *Nature*, vol. 631, pp. 755–759, 2024.

[3] A. Sriram, J. Zbontar, T. Murrell, A. Defazio, C. L. Zitnick, *et al.*, "End-to-end variational networks for accelerated mri reconstruction," in *MICCAI*, 2020, pp. 64–73.

[4] **J. Wang**, O. Ostras, B. Tolooshams, M. Sode, Z. Li, *et al.*, "Lung histology reconstruction from ultrasound radio-frequency signal via physics simulation," *Nature (in submission)*, 2024.

[5] **J. Wang**, Y. Aborahama, J. Berner, Z. Li, K. Azizzadenesheli, *et al.*, "Fast and resolution-invariant 3d photoacoustic computed tomography via operator learning," *Nature Communications (in submission)*, 2024.

[6] B. Tolooshams, **J. Wang**, L. Lin, T. C. K. Azizzadenesheli, R. A. Andersen, *et al.*, "Temporal neural operator for fast functional ultrasound imaging enabling real-time brain-computer interface," *Nature Communications (in submission)*, 2024.

[7] Z. Liu, Z. Miao, X. Zhan, **J. Wang**, B. Gong, *et al.*, "Large-scale long-tailed recognition in an open world," in *CVPR*, 2019, pp. 2537–2546.

[8] Z. Liu, Z. Miao, X. Zhan, **J. Wang**, B. Gong, *et al.*, "Open long-tailed recognition in a dynamic world," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022.

[9] Z. Miao, K. M. Gaynor, **J. Wang**, Z. Liu, O. Muellerklein, *et al.*, "Insights and approaches using deep learning to classify wildlife," *Scientific reports*, vol. 9, no. 1, pp. 1–9, 2019.

[10] **J. Wang**, T. N. Yeh, R. Chakraborty, X. Y. Stella, and M. C. Lin, "A deep learning approach for meibomian gland atrophy evaluation in meibography images," *Translational vision science & technology*, vol. 8, no. 6, pp. 37–37, 2019.

[11] **J. Wang**, Y. Chen, and S. X. Yu, "Pose-Aware Self-Supervised Learning with Viewpoint Trajectory Regularization," in *ECCV*, 2024.

[12] C.-H. Yeh, **J. Wang**, A. D. Graham, A. J. Liu, B. Tan, *et al.*, "Insight: A multi-modal diagnostic pipeline using llms for ocular surface disease diagnosis," in *MICCAI*, 2024.

[13] R. Kocielnik, E. Y. Wong, T. N. Chu, L. Lin, **J. Wang**, *et al.*, "Deep multimodal fusion for surgical feedback classification," in *Machine Learning for Health (ML4H)*, PMLR, 2023, pp. 256–267.

[14] **J. Wang**, S. Li, T. N. Yeh, R. Chakraborty, A. D. Graham, *et al.*, "Quantifying meibomian gland morphology using artificial intelligence," *Optometry and Vision Science*, vol. 98, no. 9, pp. 1094–1103, 2021.

[15] **J. Wang**, A. D. Graham, S. X. Yu, and M. C. Lin, "Predicting demographics from meibography using deep learning," *Scientific Reports*, vol. 12, no. 1, pp. 1–11, 2022.

[16] A. D. Graham, T. Kothapalli, **J. Wang**, J. Ding, V. Tse, *et al.*, "A machine learning approach to predicting dry eye-related signs, symptoms and diagnoses," *Heliyon*, 2024.

[17] A. D. Graham, **J. Wang**, T. Kothapalli, J. Ding, H. Tasho, *et al.*, "Artificial intelligence models utilize lifestyle factors to predict dry eye related outcomes," *Scientific reports*, 2024.

[18] **J. Wang**, Y. Chen, R. Chakraborty, and S. X. Yu, "Orthogonal convolutional neural networks," in *CVPR*, 2020, pp. 11 505–11 515.

[19] **J. Wang**, Y. Chen, S. X. Yu, B. Cheung, and Y. LeCun, "Compact and optimal deep learning with recurrent parameter generators," in *WACV*, 2023.