

Statement of Purpose for Declan Kutscher

Vision models for wildfire smoke detection frequently misclassify fog and clouds as smoke due to misinterpretation of spectral correlation as a physical cause. In wildfire emergency response systems, such errors can delay or misdirect firefighting and evacuation efforts, highlighting a critical limitation. Although many vision models perform well on benchmarks, they often fail to capture the spatial and physical structure present in real-world data. This limitation has significant consequences in domains such as climate science and disaster response, where model performance directly influences decision-making and public safety. **My research objective is to develop self-supervised systems that dynamically allocate computation, memory, and representation according to task and input complexity, drawing on neuroscience-inspired mechanisms.** By learning to focus computational resources and adapt representations, these systems can achieve robust performance under real-world resource constraints.

Physics Guided Learning:

At the University of Pittsburgh, with Dr. Xiaowei Jia, I investigated how physical structure can guide machine learning. In our *Physics-Guided Fair Graph Sampling* project (**He et al. AAAI 2025**), thermodynamic heat-transfer equations were embedded into a graph neural network to predict water temperature across large watersheds. The resulting model **achieved a root mean squared error (RMSE) of 1.74 compared to 1.77 for GraphSAGE** (Hamilton et al., 2018), **while reducing fairness disparity by 35%** across socioeconomic subgroups. Aligning learning algorithms with physical principles improved both reliability and equity, particularly in underrepresented regions. This work demonstrated that incorporating domain structure makes models not only more reliable but also more generalizable, a principle that continues to shape my approach to adaptive representation learning.

Representation and Efficiency Tradeoffs:

My master's thesis expanded the focus on structure and adaptation by comparing *ImageNet* (Russakovsky et al., 2015) and satellite-pretrained models. ImageNet features were found to be richer and more separable, achieving H-scores of 13 compared to 5.9. Cluster separability exceeded 1300 for ImageNet features but fell below 700 for satellite-pretrained models, indicating that **richer representations support broader generalization**. However, satellite-pretrained models **achieved higher F1 scores in small-data regimes** (0.70 versus 0.65 at 10% of *BigEarthNet*, Sumbul et al., 2019), demonstrating greater sample efficiency. *These results highlight a tradeoff between representational richness and domain specificity.* Consequently, I pivoted to self-supervised learning, seeking to acquire inductive biases directly from data rather than imposing them manually.

Adaptive Perception and Attention:

At **Berkeley AI Research (BAIR)**, I began addressing how perception order shapes visual learning through *REOrdering Patches Improves Vision Models* (**Kutscher et al. NeurIPS 2025**). We used reinforcement learning to optimize the sequence of image patches processed by transformers. Aligning input order with model bias improved performance without architectural modifications: **accuracy increased from 49.2 to 62.6 percent on *Functional Map of the World*** (Christie et al., 2018) and from **55.2 to 58.3 percent on ImageNet**. These results indicate that controlling perception order alone can enhance learning. Building on this, my PhD research will, in part, extend REOrder to determine **where to focus computation, not just the order of processing**. This approach will employ reinforcement learning policies over windowed attention patterns, token subsets, or conditioning mechanisms, enabling the model to **prune uninformative regions and allocate computation adaptively based on task and sample difficulty**. The primary challenge is to efficiently search the vast space of possible orderings while learning differentiable control signals for pruning and conditioning. Selective allocation of attention, analogous to human foveation, is expected to improve model efficiency and adaptability to structural and task demands.

Embodiment as an Efficiency Constraint:

Biological systems **develop complex skills through intrinsically motivated exploration**, creating efficient representations without external supervision. At **Carnegie Mellon University's NeuroAgents Lab**, I investigate this process by developing self-supervised reinforcement learning agents that acquire composable motor primitives via world models. While this work focuses on the mechanism of skill discovery, it highlights a fundamental constraint: to emulate biological agility, embodied learning must be computationally efficient. This makes the robotic platform the ideal testbed for my proposed research. My work on adaptive attention aims to enhance intrinsic learning by pruning uninformative visual data, allowing agents to construct world models with minimal computational waste. Just as humans naturally filter stimuli to learn skills efficiently, I aim to equip agents with visual systems that learn where to look to support autonomous, resource-constrained skill acquisition.

Toward Adaptive Self-Supervised Vision:

From physics-guided modeling to representational analysis, selective attention, and intrinsically motivated exploration, each has refined my understanding of how structure from environment, data, or perception can be leveraged to make models both efficient and general. My PhD research will unify these principles into a framework for **adaptive self-supervised vision that learns to control computation through feedback between perception, memory, and task context**. The central question is when to prune computation versus when to allocate more capacity, balancing accuracy against cost without explicit supervision. Addressing this requires developing reward signals that capture both task performance and efficiency. *This work aims to build systems that move beyond static perception toward real-time, resource-aware understanding in domains such as climate monitoring, disaster response, and embodied intelligence.*

The **University of Maryland, College Park** offers an environment where my research can thrive. I hope to work with **Dr. Ritwik Gupta**, whose work on physics-integrated computer vision and adaptive sensing for humanitarian assistance and disaster response directly connects to my goal of developing resource-aware vision systems for high-stakes, real-world deployment. Dr. Gupta has been a long time mentor and inspiration for my work as well as a co-author on my **REOrder** paper. I hope to extend our collaboration toward adaptive attention control for perception under uncertainty. **Dr. Ruoshi Liu's** research on learning visual priors and physically grounded structure in large vision models aligns with my objective to integrate environmental and spatial structure into self-supervised representations that generalize across domains. Finally, **Dr. Yiannis Aloimonos's** work on neuromorphic sensing and neuro-dynamical systems complements my interest in neuroscience-inspired control of perception and computation. Together, these faculty form an ideal environment to unify physics-guided learning, adaptive self-supervised vision, and neuro-inspired efficiency into scalable, embodied intelligence.

I am committed to mentorship and building inclusive research communities. As a teaching assistant, I developed skills in communicating complex technical concepts and supporting students from diverse backgrounds. With interdisciplinary training spanning neuroscience-inspired vision, climate science, and disaster response, I am prepared to contribute technical advances while fostering collaborative research environments. Long-term, I aim to build a research group that bridges computer vision and neuroscience to develop adaptive systems with real-world impact, grounded in rigor, openness, and mentorship.

References:

E. He, **D. Kutscher**, Y. Xie, J. Zwart, Z. Jiang, H. Yao, and X. Jia, (2025). Physics-Guided Fair Graph Sampling for Water Temperature Prediction in River Networks. Proceedings of the AAAI Conference on Artificial Intelligence, 39(27), 28070-28078.
<https://doi.org/10.1609/aaai.v39i27.35025>

- D. Kutscher**, D. M. Chan, Y. Bai, T. Darrell, and R. Gupta, "REOrdering Patches Improves Vision Models," arXiv preprint arXiv:2505.23751, 2025. [Online]. Available: <https://arxiv.org/abs/2505.23751>
- G. Christie, N. Fendley, J. Wilson, and R. Mukherjee, "Functional Map of the World." 2018. [Online]. Available: <https://arxiv.org/abs/1711.07846>
- O. Russakovsky et al., "ImageNet Large Scale Visual Recognition Challenge." 2015. [Online]. Available: <https://arxiv.org/abs/1409.0575>
- G. Sumbul, M. Charfuelan, B. Demir, and V. Markl, "Bigearthnet: A Large-Scale Benchmark Archive for Remote Sensing Image Understanding," Jul. 2019. doi: 10.1109/igarss.2019.8900532.
- W. L. Hamilton, R. Ying, and J. Leskovec, "Inductive Representation Learning on Large Graphs." 2018. [Online]. Available: <https://arxiv.org/abs/1706.02216>