Advancing Out-of-Distribution Detection Across Diverse Scenarios

Xiang Fang

Interdisciplinary Graduate Programme, Nanyang Technological University, Singapore
xiang003@e.ntu.edu.sg

Abstract

The safe deployment of artificial intelligence systems hinges on their ability to recognize and appropriately handle inputs they have not been trained for. Out-of-Distribution (OOD) detection aims to provide this capability, yet most existing methods are developed under idealized assumptions that do not hold in the real world. This thesis challenges these assumptions by systematically addressing four key practical challenges: the semantic ambiguity of unlabeled data, the presence of domain shifts and class imbalances, the scarcity of labeled training data, and the need to operate on dynamic video streams instead of static images. The core of this research is a suite of four novel deep learning frameworks, each designed to overcome one of these specific limitations. My contributions push the field of OOD detection from a laboratory problem towards a robust and practical technology, essential for building trustworthy AI.

Introduction and Motivation

Deep neural networks have achieved remarkable success in closed-set environments where the test data follows the same distribution as the training data (Fang et al. 2025a, 2023b, 2022, 2023a; Fang, Fang, and Wang 2025; Fang, Easwaran, and Genest 2025; Fang and Fang 2026; Fang, Fang, and Wang 2026; Fang et al. 2026, 2025c, 2024b, 2025b, 2024a,c, 2021b; Fang, Easwaran, and Genest 2025; Fang et al. 2020, 2021a; Fang, Easwaran, and Genest 2024). However, in real-world applications such as autonomous driving or medical diagnosis, a model will inevitably encounter novel or unexpected inputs. A standard classifier will forcedly, and often with high confidence, misclassify such inputs, with potentially catastrophic consequences. Out-of-Distribution (OOD) detection is the critical task of identifying these novel inputs at test time. While the field has made significant progress, the majority of research operates under a set of convenient yet flawed assumptions: (1) an abundance of well-labeled in-distribution (ID) data is available for training; (2) training and test data are drawn from the same well-behaved, class-balanced domain; and (3) the problem is confined to static images.

My thesis argues that for OOD detection to become a reliable component of real-world AI systems, these idealized assumptions must be systematically dismantled and addressed. My research investigates the following central question: *How can we design OOD detection methods that are robust to the complexities and constraints of real-world data, including semantic ambiguity, data imperfections, data scarcity, and dynamic modalities?*

Proposed Research Plan and Contributions

My research plan is structured around addressing four distinct, practical challenges in OOD detection. The core technical work for each of these challenges has been completed and published in peer-reviewed conferences and journals. This work was conducted under the guidance of my supervisors, Arvind Easwaran, Blaise Genest, and P. N. Suganthan, and in collaboration with my co-authors on the respective publications.

Contribution 1: Handling Semantic Ambiguity in Unlabeled Data. A common assumption is that any unlabeled data used during training is purely OOD. I challenge this by addressing the problem of **multi-granularity semantics**, where unlabeled data may contain ID samples at a finer level of detail (e.g., training on "cat" and finding "tabby cat" in the unlabeled set). **Contribution:** I developed the Adaptive Hierarchical Graph Cut (AHGC) network (Fang, Easwaran, and Genest 2025), the first graph-based method for this problem. AHGC constructs a k-NN graph over all samples and uses a hierarchical cutting mechanism to cluster semantically similar images, regardless of their initial labels. This allows it to discover and correctly identify fine-grained ID samples within a pool of unlabeled data.

Contribution 2: Tackling Data Imperfections. Real-world data is often imperfect, exhibiting both domain shifts (e.g., training on cartoons, testing on photos) and severe class imbalances. I address this combined challenge under the new problem setting of **Class-imbalanced Cross-Domain OOD Detection (CCOD)**. **Contribution:** I proposed the Uncertainty-aware Adaptive Semantic Alignment (UASA) network (Fang et al. 2025a). UASA uses a novel prototype-based alignment strategy. It builds robust, label-driven prototypes from the source domain and uses them to guide classification in the target domain. It handles the domain, semantic, and class-imbalance gaps simultaneously through adaptive thresholding and uncertainty-aware clustering.

Contribution 3: Overcoming Data Scarcity. The requirement for large labeled datasets is a major bottleneck for OOD detection. I tackle this through the problem of **Few-Shot OOD Detection**, where the model must learn from only a handful of examples per class. **Contribution:** I designed the Adaptive Multi-prompt Contrastive Network (AMCN) (Fang, Easwaran, and Genest 2025), which leverages the power of pre-trained vision-language models like CLIP. Instead of just learning prompts for ID classes, I introduced the concept of learnable OOD prompts. By explicitly creating a contrastive objective between ID and OOD prompts, AMCN learns a robust decision boundary even with extreme data scarcity.

Contribution 4: Expanding to Dynamic Modalities. The final frontier I explore is moving beyond static images to dynamic video streams. I introduced the novel task of **OOD Action Detection (ODAD)**, which requires localizing and rejecting unknown actions in untrimmed videos. **Contribution:** I created the Uncertainty-Guided Appearance-Motion Association Network (UAAN) (Fang, Easwaran, and Genest 2024). This framework uses separate branches to model appearance and motion features, reasoning over their interaction using spatial-temporal graphs. It incorporates an uncertainty estimation module to explicitly distinguish OOD actions from ID ones.

Progress and Timeline

My doctoral research is structured to culminate in a thesis that synthesizes these contributions into a unified narrative.

Progress as of September 30, 2025. The core technical contributions of my thesis are complete. The four primary research thrusts described above have each resulted in a peer-reviewed publication, demonstrating the novelty and efficacy of the proposed methods. The conceptual framework for the thesis, which ties these individual contributions together as a systematic effort to overcome the idealized assumptions of prior OOD research, is well-defined.

Anticipated Progress by January 20-21, 2026. My focus between now and the Doctoral Consortium is on the synthesis and completion of the dissertation document itself. The timeline is as follows: 1) **October - November 2025:** Complete the final drafts of the introduction, background, and related work chapters. Write the concluding chapter, which will synthesize the four contributions and discuss their collective impact and future research directions. 2) **December 2025:** A complete draft of the entire thesis will be assembled and submitted to my supervisory committee for initial feedback. 3) **January 2026:** Based on feedback, I will have a polished, full draft of my thesis. The presentation at the Doctoral Consortium will be based on this completed narrative, allowing me to share the full story of my research and receive valuable feedback before my final defense.

Anticipated Thesis Contribution

The primary contribution of this thesis is a significant push towards making OOD detection more practical and applicable to real-world AI systems. By moving beyond the field's traditional, idealized assumptions, my work provides a suite of solutions for handling the complex and data-scarce environments where AI is increasingly deployed. The thesis will not only present four novel methods but will also provide a new perspective on how to approach OOD research: by identifying and directly tackling the practical constraints that limit the deployment of otherwise effective algorithms.

References

Fang, X.; Easwaran, A.; and Genest, B. 2024. Uncertainty-Guided Appearance-Motion Association Network for Out-of-Distribution Action Detection. In *MIPR*.

Fang, X.; Easwaran, A.; and Genest, B. 2025. Adaptive Multi-prompt Contrastive Network for Few-shot Out-of-distribution Detection. In *ICML*.

Fang, X.; Easwaran, A.; Genest, B.; and Suganthan, P. N. 2025a. Your data is not perfect: Towards cross-domain out-of-distribution detection in class-imbalanced data. *ESWA*.

Fang, X.; and Fang, W. 2026. Disentangling Adversarial Prompts: A Semantic-Graph Defense for Robust LLM Security. In *AAAI*.

Fang, X.; Fang, W.; Ji, W.; and Chua, T.-S. 2025b. Turing Patterns for Multimedia: Reaction-Diffusion Multi-Modal Fusion for Language-Guided Video Moment Retrieval. In *MM*.

Fang, X.; Fang, W.; Liu, D.; Qu, X.; Dong, J.; Zhou, P.; Li, R.; Xu, Z.; Chen, L.; Zheng, P.; et al. 2024a. Not all inputs are valid: Towards open-set video moment retrieval using language. In *MM*.

Fang, X.; Fang, W.; and Wang, C. 2025. Hierarchical Semantic-Augmented Navigation: Optimal Transport and Graph-Driven Reasoning for Vision-Language Navigation. In *NeurIPS*.

Fang, X.; Fang, W.; and Wang, C. 2026. Unveiling the Fragility of Vision-Language Models: Multi-Modal Adversarial Synergy via Texture-Constrained Perturbations and Cross-Modal Optimization. In *AAAI*.

Fang, X.; Fang, W.; Wang, C.; Liu, D.; Tang, K.; Dong, J.; Zhou, P.; and Li, B. 2025c. Multi-pair temporal sentence grounding via multi-thread knowledge transfer network. In *AAAI*.

Fang, X.; Fang, W.; Wang, C.; Qu, X.; and Liu, D. 2026. Rethinking Video-language Model From the Language Input Perspective. In *AAAI*.

Fang, X.; Hu, Y.; Zhou, P.; and Wu, D. 2021a. Animc: A soft approach for autoweighted noisy and incomplete multiview clustering. *TAI*.

Fang, X.; Hu, Y.; Zhou, P.; and Wu, D. O. 2020. V3H: View variation and view heredity for incomplete multiview clustering. *TAI*.

Fang, X.; Hu, Y.; Zhou, P.; and Wu, D. O. 2021b. Unbalanced incomplete multi-view clustering via the scheme of view evolution: Weak views are meat; strong views do eat. *TETCI*.

Fang, X.; Liu, D.; Fang, W.; Zhou, P.; Xu, Z.; Xu, W.; Chen, J.; and Li, R. 2024b. Fewer Steps, Better Performance: Efficient Cross-Modal Clip Trimming for Video Moment Retrieval Using Language. In *AAAI*.

Fang, X.; Liu, D.; Zhou, P.; and Hu, Y. 2022. Multi-modal cross-domain alignment network for video moment retrieval. *TMM*.

Fang, X.; Liu, D.; Zhou, P.; and Nan, G. 2023a. You can ground earlier than see: An effective and efficient pipeline for temporal sentence grounding in compressed videos. In *CVPR*.

Fang, X.; Liu, D.; Zhou, P.; Xu, Z.; and Li, R. 2023b. Hierarchical local-global transformer for temporal sentence grounding. *TMM*.

Fang, X.; Xiong, Z.; Fang, W.; Qu, X.; Chen, C.; Dong, J.; Tang, K.; Zhou, P.; Cheng, Y.; and Liu, D. 2024c. Rethinking Weakly-supervised Video Temporal Grounding From a Game Perspective. In *ECCV*. Springer.