Capability Statement:

IARPA Video LINCS Program

Dartmouth College

1. Background

We, a group of professors from Dartmouth College, specializing in computer vision and artificial intelligence research, express our keen interest in joining and contributing to the Video LINCS program. Our team brings together a wealth of experience and expertise in the fields of **object detection**, **tracking**, **re-identification**, **deep learning**, and **open set classification**.

2. Co-PI Information

Yu-Wing Tai: Currently an Associate Professor in the Computer Science Department at Dartmouth College. Prior to joining Dartmouth, Yu-Wing Tai was a senior research director at Kuaishou Technology in China (2020 to 2023), a research director at Tencent (2017 to 2020), a principal researcher at SenseTime (2015-2016), and an associate professor at KAIST, Korea (2009-2014). Yu-Wing Tai has rich experiences in both Academic and Industry, having published more than 100 papers in areas covering low-level vision, image/video editing, segmentation, 3D reconstruction, object detection, human pose estimation, and video analysis. Alongside his academic achievements, his extensive industrial background equips him to contribute to product-level development and management. The Co-PI's expertise aligns seamlessly with the objectives of the proposed project.

- Personal Homepage: https://yuwingtai.github.io/
- Google Scholar: https://scholar.google.com/citations?hl=en&user=nFhLmFkAAAAJ

Yaoqing Yang: Yaoqing Yang currently holds the position of Assistant Professor in the Department of Computer Science at Dartmouth College. Before that, he was a postdoctoral researcher at UC Berkeley. He has been working on improving machine learning models' generalizability, transparency, and robustness. His two papers on 3D point clouds have got more than 1500 citations in total. His current focus is to diagnose failures of machine learning models using geometric features from high dimensions, such as loss landscapes, spectral densities of weight matrices, and decision boundaries. He also applies these techniques to 3D point clouds and graph neural networks.

- Personal Homepage: https://sites.google.com/site/yangyaoqingcmu/
- Google Scholar: https://scholar.google.com/citations?user=LYvugWgAAAAJ

SouYoung Jin: SouYoung Jin is currently an Assistant Professor in the Department of Computer Science at Dartmouth College. Before joining Dartmouth, she was a postdoctoral associate at MIT (2020-2022). Her primary research areas include computer vision, machine learning, and cognitive science, with a specialization in video understanding. SouYoung has contributed to various research projects involving object detection, tracking, re-identification, and deep learning. During her PhD

studies, she concentrated on object detection and clustering in videos, and her work was specifically funded by IARPA.

- Personal Homepage: https://souyoungjin.github.io/
- Google Scholar: https://scholar.google.com/citations?user=_B-_CzYAAAAJ&hl=en

Yujun Yan: Yujun Yan holds the position of Assistant Professor in the Department of Computer Science at Dartmouth College. Specializing in graph learning, her research delves into the expressiveness, generalizability, and explainability of graph neural networks. Notably, her contributions include two well-received papers on learning heterophily graphs, which effectively model various types of interactions between entities. The two papers have received more than 800 citations in total. Currently, Yujun Yan is actively engaged in research aimed at addressing out-of-distribution shifts in graphs. In this project, Yujun will utilize her expertise in graph learning to extract static and dynamic person signature for person re-identification.

- Personal Homepage: https://sites.google.com/umich.edu/yujunyan/home
- Google Scholar: https://scholar.google.com/citations?user=5TQUP58AAAAJ

3. Core Competencies

3.1. Object Detection

The extensive publication record of the Co-PI, Yu-Wing Tai, reflects a profound expertise in object detection, making it particularly well-suited for addressing the challenges posed by the Video LINCS program. Notably, the paper "Towards robust object detection invariant to real-world domain shifts" presented at ICLR 2023 underscores the Co-PI's commitment to overcoming domain shift problems, showcasing a dedication to ensuring the robustness of object detection across real-world variations. Furthermore, the paper "Cascade-DETR: Delving into High-Quality Universal Object Detection," featured at ICCV 2023, highlights the PI's exploration of high-quality universal object detection, aligning with the program's goal of autonomously re-identifying objects across diverse video sensor collections. The inclusion of "Few-Shot Object Detection with Model Calibration" and "Few-shot video object detection" in ECCV 2022 emphasizes the PI's proficiency in addressing few-shot settings.

The Co-PI, SouYoung Jin, focused on automatically generating pseudo-labels by leveraging spatiotemporal relationships between objects in videos to enhance the performance of object detection, as demonstrated on her hard example mining paper in ECCV 2018.

With a proven track record in handling real-world domain shifts, few-shot scenarios, and incorporating 3D detection frameworks, the PI's expertise is poised to make significant contributions to the success of the Video LINCS program.

★ Related Representative Publications:

- Towards robust object detection invariant to real-world domain shifts, ICLR 2023
- Cascade-DETR: Delving into High-Quality Universal Object Detection, ICCV 2023

- NeRF-RPN: A general framework for object detection in NeRFs, CVPR 2023
- Few-Shot Object Detection with Model Calibration, ECCV 2022
- Few-shot video object detection, ECCV 2022
- Group collaborative learning for co-salient object detection, CVPR 2021
- Few-shot object detection with attention-RPN and multi-relation detector, CVPR 2020
- Unsupervised hard example mining from videos for improved object detection, ECCV 2018
- Automatic adaptation of object detectors to new domains using self-training. CVPR 2019

3.2. Tracking

The Co-PI, Yu-Wing Tai, demonstrates a robust capability in tracking methodologies, particularly in the context of the Video LINCS program's objectives. Noteworthy publications such as "Segment anything meets point tracking" underscore the PI's commitment to advancing tracking techniques, specifically integrating segmentation to enhance precision. The paper "Mask-free video instance segmentation" presented at CVPR 2023 showcases the PI's dedication to developing segmentation and tracking approaches without using mask data in training. Additionally, the paper "Prototypical cross-attention networks for multiple object tracking and segmentation" and "Rethinking space-time networks with improved memory coverage for efficient video object segmentation" in NeurIPS 2021 highlights the PI's expertise in integrating multiple object tracking with segmentation.

With a demonstrated proficiency in tracking methodologies, particularly with an emphasis on segmentation techniques, the PI is well-positioned to contribute significantly to the success of the Video LINCS program, addressing the critical need for autonomously tracking and segmenting objects across diverse video sensor collections.

★ Related Representative Publications:

- Segment anything meets point tracking, ArXiv 2023
- Mask-free video instance segmentation, CVPR 2023
- Video mask transfiner for high-quality video instance segmentation, CVPR 2022
- Prototypical cross-attention networks for multiple object tracking and segmentation, NeurIPS 2021
- Rethinking space-time networks with improved memory coverage for efficient video object segmentation, NeurIPS 2021

3.3. Re-identification

The Co-PI, Yu-Wing Tai, exhibits a notable expertise in re-identification, aligning seamlessly with the objectives of the Video LINCS program. The paper "Unsupervised multi-view object segmentation using radiance field propagation" (NeurIPS 2022) and "Dense hybrid recurrent multi-view stereo net with dynamic consistency checking" at ECCV 2020 underscores the PI's commitment to advancing re-identification techniques by introducing unsupervised multi-view object segmentation. The paper "Push for center learning via orthogonalization and subspace masking for person re-identification" published in IEEE Trans. on Image Processing in 2020 emphasizes the PI's proficiency in person re-identification.

The Co-PI, SouYoung Jin, proposed a clustering approach called the Erdos-Renyi clustering algorithm, which was spotlighted at ICCV 2017. This work demonstrated that the algorithm can be applied to movie clips, successfully re-identifying the same characters in videos.

★ Related Representative Publications:

- Unsupervised multi-view object segmentation using radiance field propagation, NeurIPS 2022
- Dense hybrid recurrent multi-view stereo net with dynamic consistency checking, ECCV 2020
- Push for center learning via orthogonalization and subspace masking for person re-identification, TIP, 2020
- End-to-end face detection and cast grouping in movies using Erdos-Renyi clustering, ICCV 2017

3.4. Deep Learning

The Co-PI Yaoqing Yang's contributions to the Video LINCS program in the deep learning direction consists of the following directions: (1) Spatial reasoning neural networks. Spatial reasoning can be a pretraining task to provide a strong prior for video object detection networks. Yang's expertise in spatial reasoning is highlighted by the CVPR 2021 paper "Self-Supervised Spatial Reasoning on Multi-View Line Drawings." (2) 3D point cloud neural networks. Yang has worked extensively on point cloud deep learning, which can be seen in his CVPR 2018 paper "Foldingnet: Point cloud auto-encoder via deep grid deformation" and "Mining Point Cloud Local Structures by Kernel Correlation and Graph Pooling," as well as his 2019 paper "Deep unsupervised learning of 3D point clouds via graph topology inference and filtering" published in Transactions on Image Processing.

Co-PI SouYoung Jin has applied deep learning in her recent papers on video understanding, particularly exploring multiple modalities, including videos, audios, and text. For example, in her ACL 2021 paper "Cross-Modal Discrete Representation Learning", concepts have been trained across different modalities. In addition, her recent work focuses on leveraging LLMs by translating audiovisual features into representations understandable by LLMs.

Co-PI Yujun will contribute to the Video LINCS program by exploring effective graph modeling and augmentations. (1) Effective graph modeling: Yujun has rich experience in modeling diverse types of interactions between entities, shown in her NeurIPS 2020 paper "Beyond Homophily in Graph Neural Networks: Current Limitations and Effective Designs." and her ICDM 2022 paper "Two Sides of the Same Coin: Heterophily and Oversmoothing in Graph Convolutional Neural Networks." (2) Augmentations: Yujun has experience in exploring domain-aware graph augmentations, which is illustrated in her WWW 2022 paper "Augmentations in graph contrastive learning: current methodological flaws & towards better practices".

The usage of spatial reasoning, LLMs and efficient implementation of deep learning models is crucial for real-time processing of diverse video data, and the PI's expertise positions them as a valuable asset to the success of the program.

★ Related Representative Publications:

- Self-Supervised Spatial Reasoning on Multi-View Line Drawings, CVPR 2021
- Foldingnet: Point cloud auto-encoder via deep grid deformation, CVPR 2018
- Mining Point Cloud Local Structures by Kernel Correlation and Graph Pooling, CVPR 2018

- Deep unsupervised learning of 3D point clouds via graph topology inference and filtering, TIP, 2019
- Beyond Homophily in Graph Neural Networks: Current Limitations and Effective Designs. NeurIPS 2020
- Two Sides of the Same Coin: Heterophily and Oversmoothing in Graph Convolutional Neural Networks. ICDM 2022
- Augmentations in graph contrastive learning: current methodological flaws & towards better practices. WWW 2022
- Cross-Modal Discrete Representation Learning, ACL 2022
- Leveraging Temporal Context in Low Representational Power Regimes, CVPR 2023
- LangNav: Language as a Perceptual Representation for Navigation, Arxiv 2023

3.5. Open Set Classification Learning

The Co-PI, Yu-Wing Tai, also showcases a significant contribution in the realm of open set classification. The publication "Segment Anything in High Quality," presented at NeurIPS 2023 focusing on segmenting diverse elements with high-quality masks in open world settings.

Co-PI Yaoqing Yang's contributions to the Video LINCS program in the openset classification direction consists of his studies on non-robust features and uncertainty quantification. For example, see his recent publication "Boundary thickness and robustness in learning models" in NeurIPS 2020.

Co-PI SouYoung Jin constructed a new video description dataset, the Spoken Moments, showcasing that models trained on this dataset exhibit significantly improved performance in zero-shot settings.

Co-PI Yujun Yan will contribute to the open set classification by utilizing her past experience in out-of-distribution generalization. In her NeurIPS 2020 paper "Neural Execution Engines: Learning to Execute Subroutines" and her recent preprint "Size Generalizability of Graph Neural Networks on Biological Data: Insights and Practices from the Spectral Perspective", she explored deep models capable of generalizing to unseen data sizes significantly larger than those encountered during training.

★ Related Representative Publications:

- Segment Anything in High Quality, NeurIPS 2023
- HAA500: Human-centric atomic action dataset with curated videos, CVPR 2021
- Spoken Moments: Learning Joint Audio-Visual Representations from Video Descriptions, CVPR 2021
- Boundary thickness and robustness in learning models, NeurIPS 2020
- Neural Execution Engines: Learning to Execute Subroutines, NeurIPS 2020
- Size Generalizability of Graph Neural Networks on Biological Data: Insights and Practices from the Spectral Perspective, arXiv 2023

4. Conclusion

In joining the Video LINCS program, our group of professors stands ready to contribute significantly to the development of novel capabilities in ***object detection, tracking, re-identification, deep learning, and open set classification***. Our collective expertise, collaborative ethos, and commitment to pushing technological boundaries position us as ideal contributors to the success of the program. We look forward to the opportunity to engage, collaborate, and deliver innovative solutions that exceed expectations.