**Topic:** Multimodal Learning

**Project Title:** Multimodal Learning for Detecting and Tracking Unmanned Aerial Vehicle

**Description of the Project:**

With the rapid rise in the deployment of Unmanned Aerial Vehicles (UAVs) across various sectors, there is a pressing need for robust, long-range detection and tracking systems. Traditional radar systems remain a widely used technology for aerial detection, yet they are often optimized for larger aircraft and may fall short when detecting small UAVs due to their low radar cross-section. Furthermore, these systems typically require specialized infrastructure, making them impractical for mobile or UAV-mounted deployments.LIDAR technology, with its high-resolution 3D point cloud generation, offers an effective alternative for detecting UAVs by capturing their spatial structure and movement patterns. However, LIDAR can be sensitive to atmospheric conditions and may struggle in foggy or dusty environments.RGB imaging remains a popular and cost-effective modality for UAV detection and localization, particularly under favorable lighting. However, performance tends to degrade in low-light or nighttime conditions, necessitating the integration of complementary sensing methods.Acoustic sensors offer another promising avenue by detecting the distinctive sound signatures of UAVs. While these sensors provide valuable information, especially in close to mid-range scenarios, they are limited by environmental noise and reduced effectiveness over long distances.

Given the limitations of individual sensing modalities, this project proposes a **multimodal machine learning approach** that integrates **RGB imaging, acoustic sensing, LIDAR, and radar data** to enable reliable, long-distance UAV detection and tracking under diverse environmental conditions. The fusion of these complementary data sources aims to overcome the weaknesses of each sensor alone, thereby enhancing the robustness, accuracy, and real-time applicability of the system. **We will also explore and evaluate different modality combinations to determine the most effective and efficient sensor fusion strategy.**

A key challenge in such multimodal systems lies in the proper synchronization and alignment of heterogeneous sensor data. Ensuring that each modality contributes meaningful and temporally consistent features is essential for effective model training and inference. This project emphasizes preprocessing strategies, time alignment, and feature fusion techniques to develop a unified detection and tracking framework.Ultimately, the goal is to deliver a scalable, affordable, and high-performance UAV detection system that leverages the strengths of multiple sensing modalities and advances the field of aerial surveillance and security.

**Task for the Project:**

1. **Survey existing datasets** focused on multimodal UAV detection and tracking using RGB, audio, LIDAR, and radar.
2. **Review multimodal detection and tracking methods** to understand current strategies and fusion techniques.
3. **Analyze challenges and limitations** related to small UAV detection, including sensor and environmental constraints.
4. **Select optimal sensor modalities** based on effectiveness, complementarity, and environmental adaptability.
5. **Develop unimodal detection models** to extract meaningful features from each individual modality.
6. **Implement feature fusion techniques** to integrate multimodal data for robust UAV detection and tracking.
7. **Define evaluation metrics** to assess system accuracy, speed, robustness, and suitability for real-world use.
8. **Validate system performance** using metrics and ablation tests to analyze the contribution of each modality.

**Skills Needed:**

* Basic proficiency in Python programming.
* Understanding of vector and matrix operations.

**Fundamental Materials to Begin:**

[1] Yuan, Shenghai, et al. "MMAUD: A comprehensive multi-modal anti-UAV dataset for modern miniature drone threats." *2024 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2024.

[2] Redmon, Joseph, et al. "You only look once: Unified, real-time object detection." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.

[3] **[YOLO8 and LiDAR Object Detection](https://www.youtube.com/playlist?list=PLjKiqVxeHMqnW4MoQ6rOfMN_9rh0svUPU)** (video)

[4] Cao, Yue, et al. "Multimodal object detection by channel switching and spatial attention." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2023.