C2FDrone: Coarse-to-Fine Drone-to-Drone Detection using Vision Transformer Networks

Sairam V C Rebbapragada, Pranoy Panda, Vineeth N Balasubramanian
{ai20resch13001, vineethnb}@iith.ac.in
Indian Institute of Technology, Hyderabad

ICRA2024 YOKOHAMA | JAPAN

Motivation

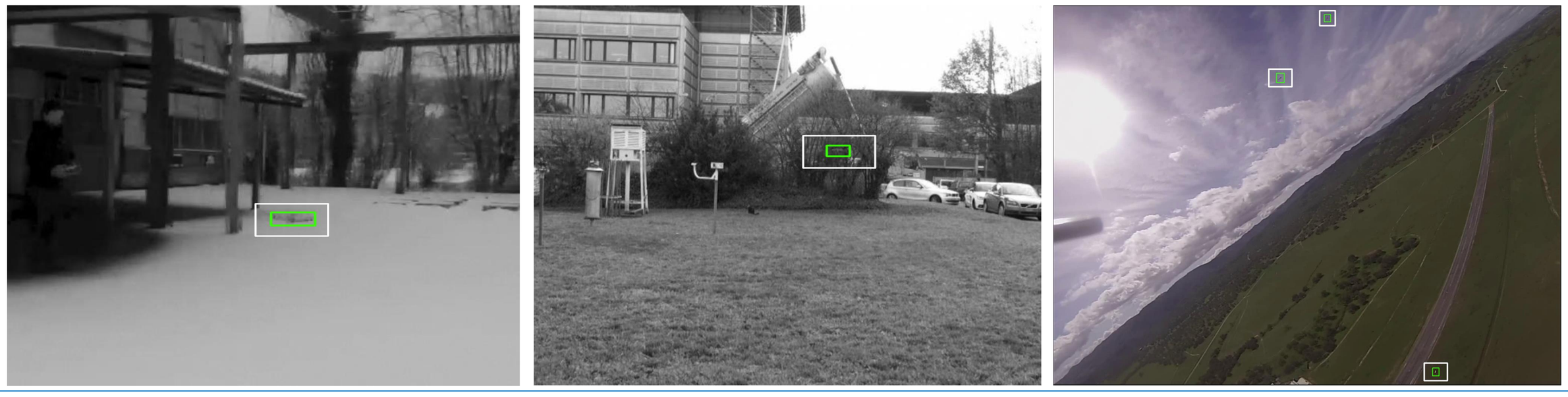
- In robotics, it is important for drones to detect each other during flight
- This helps avoid drone collisions, counter hostile drones, and facilitates drones to collaborate and cover larger areas during search-and-rescue
- While research on drone-based ground object detection has been well-studied, drone-to-drone detection remains relatively less explored.

Challenges

- Large variations in illumination levels and minimal contrast between the drones and the background
- Extremely Small Objects (Avg drone to frame ratio)

FL Drones	NPS Drones
0.07%	0.05%

What is this task about?



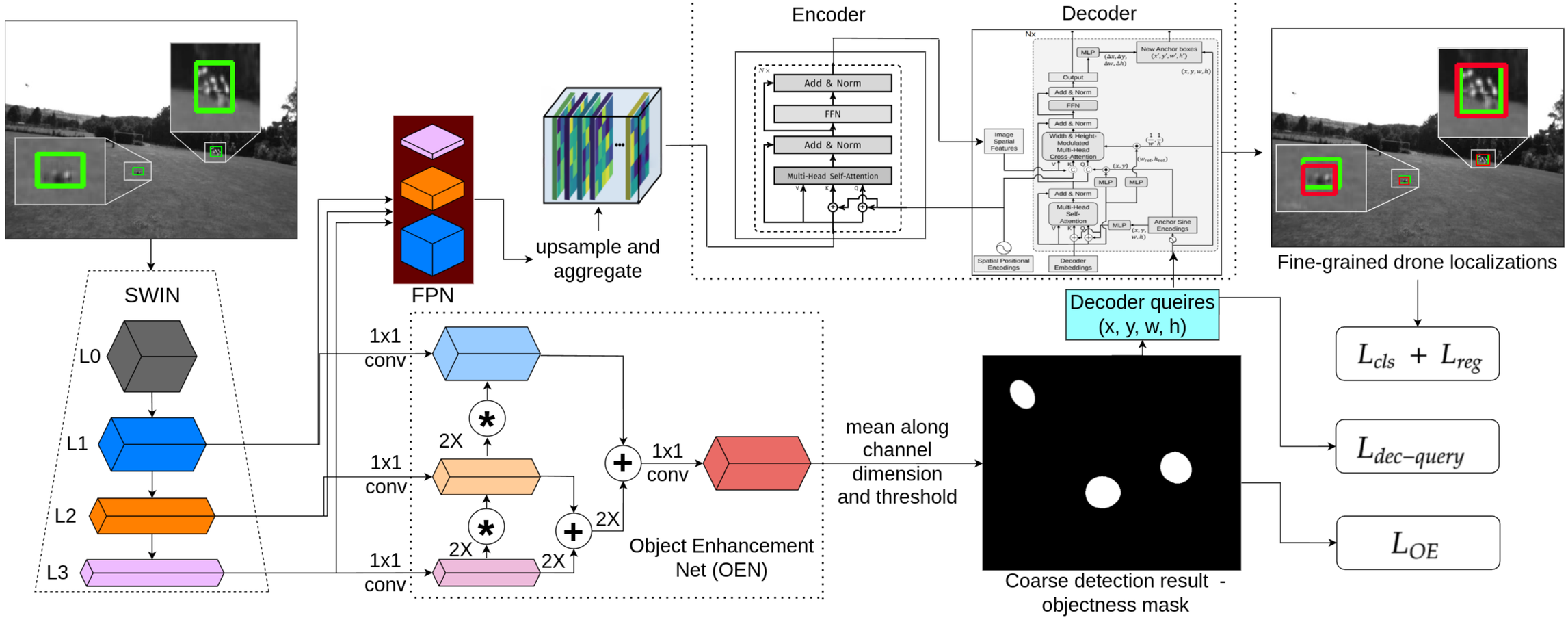
Contributions

- A novel method, **C2FDrone**, to detect and localize drones using detection transformer networks and untapped objectness information in image representations.
- C2FDrone is end-to-end trainable and works in real-time.

How can our work be useful to you?

- **For friends in Control:** For tasks such as Vision-based UAV-to-UAV pursuit our method can be a building block.
- **For friends in Perception:** Latent inductive biases in our feature extractors can be leveraged for downstream tasks.

Proposed Methodology



Loss Functions

Let ***P*** **-** Objectness mask; ***G*** - Ground truth mask

$$L_{\text{Dice}} = 1 - 2 \cdot \frac{|P \cap G|}{|P \cup G|}$$

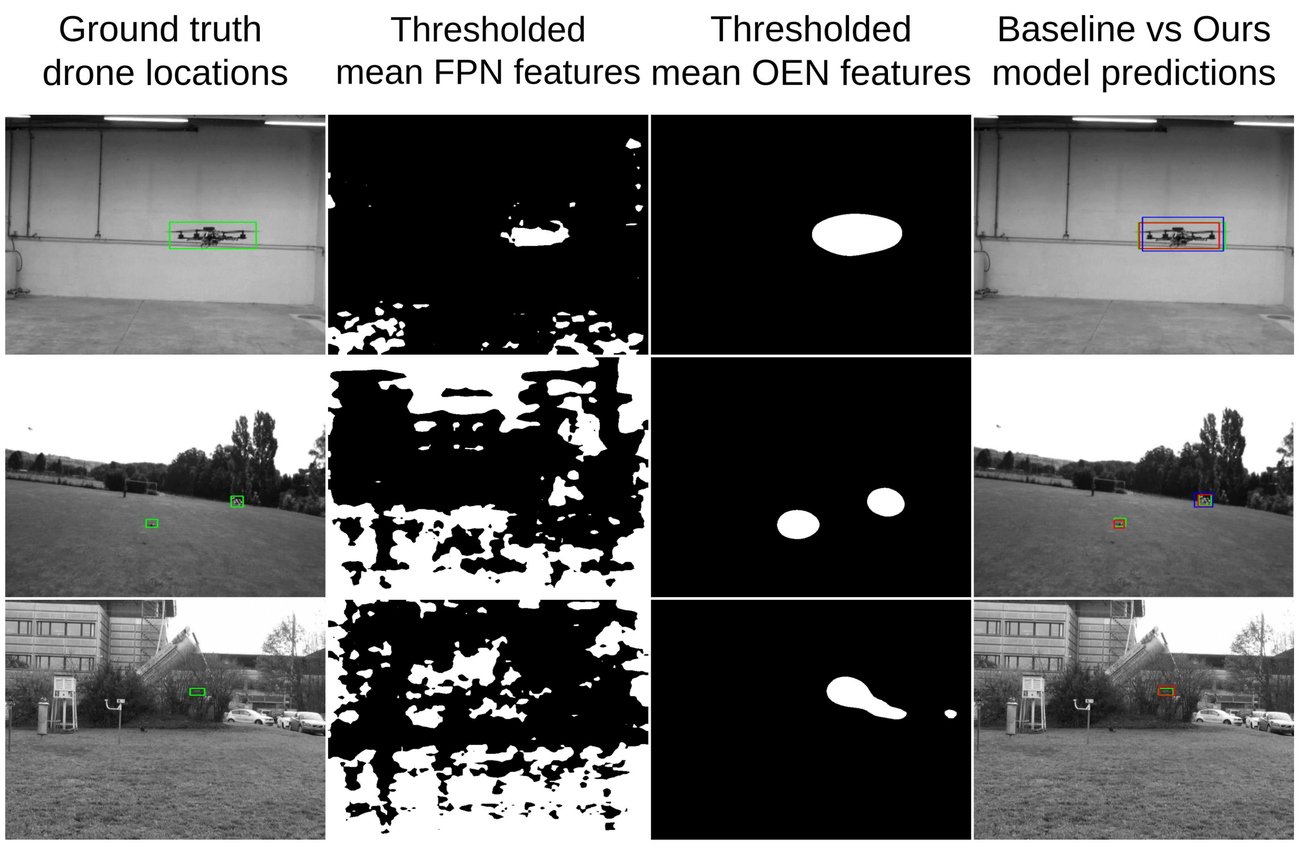
$$L_{\text{BCE}} = \sum_{i=1}^{n} -(G_i \cdot \log(P_i) + (1 - G_i) \cdot \log(1 - P_i))$$

$$L_{\text{OE}} = \alpha \cdot L_{\text{Dice}} + \beta \cdot L_{\text{BCE}}$$

Let Q and C denote the set of decoder queries (anchor boxes) and the set of highlighted regions in the objectness mask respectively.

$$L_{\text{dec-query}} = \sum_{q \in \text{Q}} \min_{c \in \text{C}} (\text{dist}(q, c))$$

Qualitative Results



Quantitative Results