

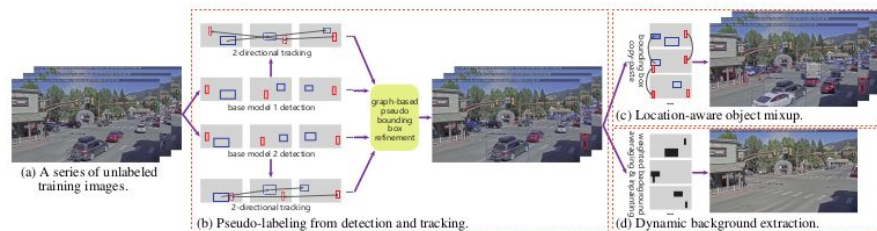
# Object Detection with Self-Supervised Scene Adaptation

## What is the problem?

The topic addresses the challenge of improving object detection in videos captured from stationary cameras, where a single object detector often fails to perform well across different scenes due to limited training data or the generalization ability of lightweight models. A key issue is adapting pre-trained object detectors to specific scenes without requiring extensive annotated data, which is costly and time-consuming.

## What has been done earlier?

Previous work in domain adaptation for object detection has explored methods like semi-supervised or weakly-supervised adaptation, self-labeling, and domain alignment. These approaches rely on generating pseudo-labels or aligning feature distributions between domains to mitigate the domain gap. However, earlier methods often struggle with noisy pseudo-labels, and domain alignment techniques are sensitive to hyperparameters and training conditions, especially when working with videos from fixed perspectives.



## **What are the remaining challenges?**

The main challenge is the difficulty of adapting a pre-trained, scene-generic detector to a specific scene without a large amount of annotated data. Previous methods face issues such as noisy pseudo-labels, difficulties in handling domain shifts, and inefficiencies in utilizing temporal data from video streams.

## **What novel solution proposed by the authors to solve the problem?**

- A cross-teaching strategy where multiple detectors work together to generate refined pseudo-labels.
- Object tracking in both forward and backward directions to improve pseudo-labels by recovering missed detections.
- A location-aware mixup technique, which minimizes artifacts in generated training data by leveraging the static nature of the background in stationary videos.
- Dynamic background modeling, which extracts background images and uses them as an additional input modality to enhance the detection accuracy.
- The introduction of a new dataset, "Scenes 100," specifically designed for testing scene-adaptive object detection.