Project EE798

Student **Bachhav Aryan Kishor**
Roll Number **210253**

Imeplementation Detais

Model: Cross-Modality Fusion Transformer (CFT) in this paper . Leveraging the self attention of the Transformer, the network can naturally carry out simultaneous intra-modality and inter-modality fusion, and robustly capture the latent interactions between RGB and Thermal domains, thereby significantly improving the performance of multispectral object detection. Extensive experiments and ablation studies on multiple datasets demonstrate that our approach is effective and achieves state-of-the-art detection performance.

Result Information:

- LLVIP Recall : 0.96
  Prescion : 0.927
  mAP 50 : 96.5
  mAP 75 : 69.6
  mAP : 60.2
  In research Paper
  mAP 50 : 97.5
  mAP 75 : 72.9
  mAP : 63.6
- Vedai mAP 50 : 73.4
  mAP : 51.6
  in research paper
  mAP 50 : 85.3
  mAP : 56

Dataset Description: LLVIP : This dataset contains 2414 images, or 1207 pairs, most of which were taken in low-light environments, and all of the images are strictly spatio-temporal aligned.Size is 640*640 Vedai : This dataset contains 2268 images, or 1134 pairs, most of which were taken in low-light environments, and all of the images are strictly spatio-temporal aligned.Size is 640*640

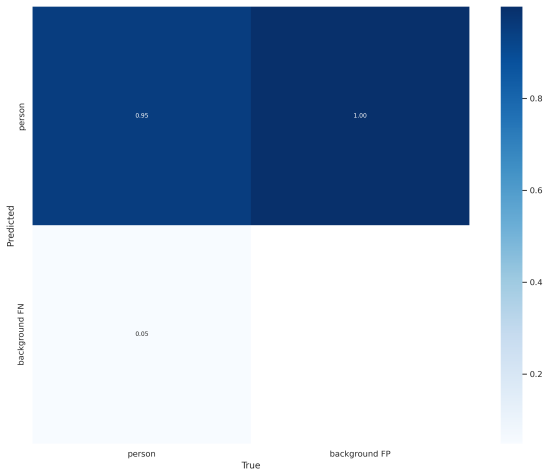

Figure 1: LLVIP Confusion Matrix

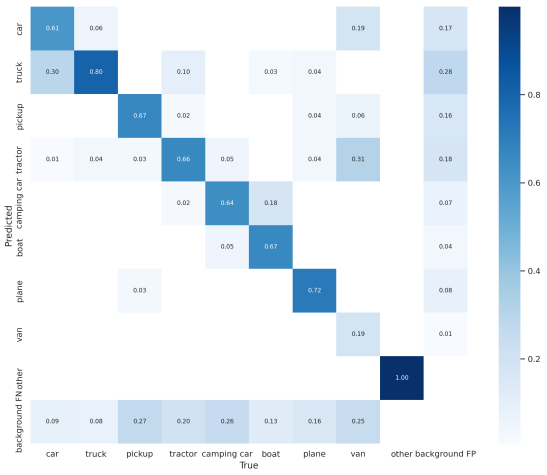

Figure 2: Vedai Confusion Matrix

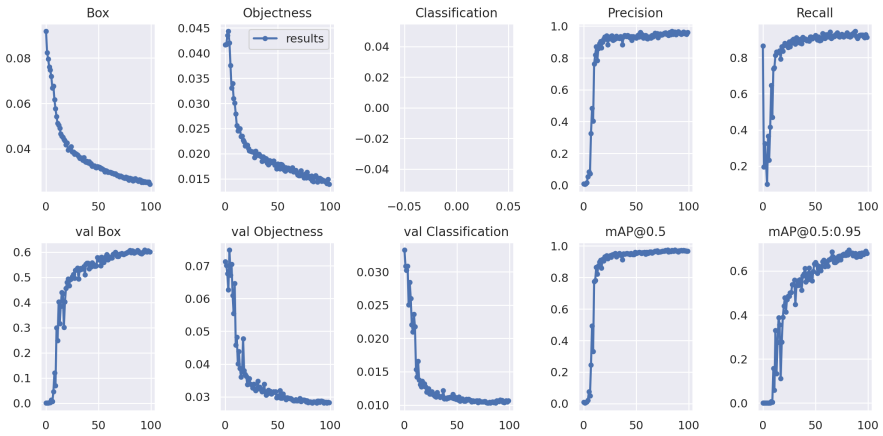

Figure 3: LLVIP Results

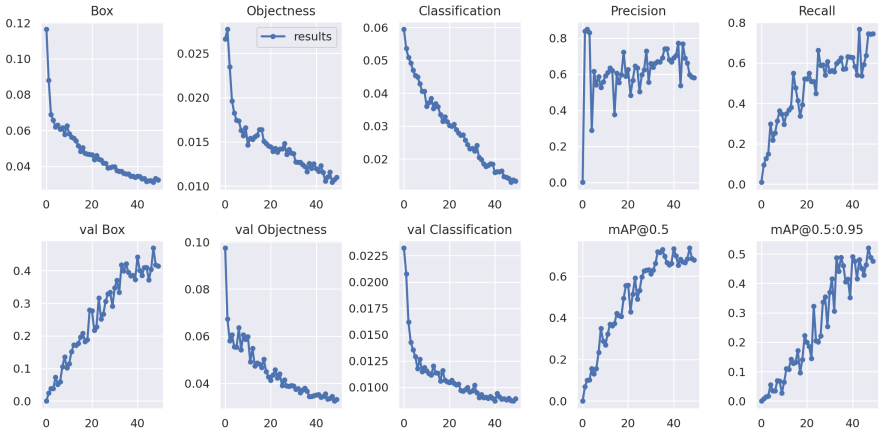

Figure 4: Vedai Result