YOLOv8++ for overlap object detection from cluttered indoor shots, invariant to sensor, lighting and affine transformation

Computer Vision (CS6350)

TPA-3

1. Problem Statement

Localization of different types of objects in indoor scenes with the help of bounding-boxes using Deep Learning techniques. The images may include challenges like occlusion, background clutter, camera shake, varied object size, affine transformation and illumination condition. The aim of the project is to detect and recognize objects in real-time, taking live feed from a web-camera (for live demo). Models based on recent state-of-the-art methods like YOLO v5[1], YOLO v3[7], RefineNet [8], Relation-Net [9], RFBNet [10], CornerNet [11] etc. or anything new showing substantial performance over the existing techniques are expected.

2. Input

Images containing objects.

3. Assumptions

Images may contain multiple objects of the same or different

classes. **4. Output**

- Detect, identify and visualize the location of each of the objects in the input image using bounding boxes
- Obtain the class label of the detected objects.
- Calculate Rank-1 recognition rate of each of the objects detected.
- Online demo taking live feed from web-camera.
- Off-line performance on a set of test images (to be provided during evaluation) using mean Average Precision (mAP@IoU=0.5) as

evaluation metric

5. Object recognition examples

6. Datasets

- *ImageNet*, *MS COCO* and *Pascal VOC* *(2011)* to be used for training followed by fine-tuning in *VPLab-Indoor* Dataset (to be provided).

7. References

1. Liu, S., Qi, L., Qin, H., Shi, J. and Jia, J., 2018. Path aggregation network for instance segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 8759-8768).

2. Jifeng Dai, Yi Li, Kaiming He, and Jian Sun. \R-FCN: Object Detection via Region-based Fully Convolutional Networks.", NIPS 2016.

3. Joseph Redmon and Ali Farhadi.\YOLO9000: Better, Faster, Stronger." CVPR 2017.

4. Tsung-Yi Lin, Piotr Dollr, Ross Girshick, Kaiming He, Bharath Hariharan, and Serge Belongie. \Feature pyramid networks for object detection." CVPR 2017.

5. Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollr. \Focal Loss for Dense Object Detection." ICCV 2017.

6. Kaiming He, Georgia Gkioxari, Piotr Dollr, and Ross Girshick. \Mask r-cnn." ICCV 2017.

7. Joseph Redmon and Ali Farhadi. \YOLOv3: An Incremental Improvement." arXiv 2018.

8. Shifeng Zhang, Longyin Wen, Xiao Bian, Zhen Lei, and Stan Z. Li. \Single-Shot Re nement Neural Network for Object Detection." CVPR 2018.

9. Han Hu, Jiayuan Gu, Zheng Zhang, Jifeng Dai, and Yichen Wei. \Relation networks for object detection." CVPR 2018.

10. Songtao Liu, and Di Huang. \Receptive eld block net for accurate and fast object detection." ECCV 2018.

11. Hei Law, and Jia Deng. \Cornernet: Detecting objects as paired keypoints." ECCV 2018.

12. Joseph, K. J., et al. "Towards open world object detection." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2021.

13. Xie, Xingxing, et al. "Oriented R-CNN for object detection." *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2021.
14. Xie, Enze, et al. "Detco: Unsupervised contrastive learning for object detection." Proceedings of the IEEE/CVF International Conference on *Computer Vision*. 2021.
15. Fan, D., Zhang, J., Xu, G., Cheng, M., & Shao, L. (2021). Salient Objects in Clutter.

**16. Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection.
In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 779-788). text**

17. https://docs.ultralytics.com/models/yolov8/#supported-tasks-and-modes