Kamali Bakthavatchalam 22081557

**Image Detection using Faster R-CNN**

**Introduction**

Image detection is a fundamental task in computer vision, enabling machines to recognize and locate objects within images. In this project, we focus on detecting people using the Faster R-CNN model, a widely used deep learning architecture known for its accuracy and speed (Ren et al., 2015). With the rise of real-time applications such as surveillance, autonomous vehicles, and smart systems, efficient person detection has become increasingly important. This work explores a dataset of person ,car, dog images, applies modern detection techniques, and evaluates the model’s ability to identify individuals across various test cases.

**Literature Review**

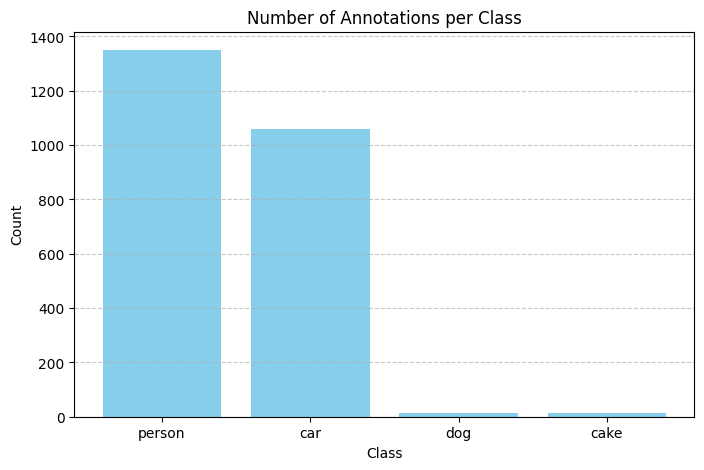
Ren et al. (2015) introduced Faster R-CNN, a two-stage object detection model that combines region proposal and classification, improving both accuracy and speed. This architecture is well-suited for tasks like person detection.

Redmon and Farhadi (2018) proposed YOLOv3, a fast one-stage detector. While it offers high speed, it may sacrifice accuracy for small or overlapping objects compared to Faster R-CNN.

Zhao et al. (2019) reviewed object detection models and confirmed that two-stage models like Faster R-CNN generally provide higher precision, making them suitable for detailed detection tasks such as identifying people in complex images.

**Data and Exploratory Data Analysis (EDA)**

The dataset used in this project is based on COCO annotations focusing on the target classes: car, person, cake, and dog. Using the COCO API, I extracted and analysed the number of annotations per class and the number of unique images containing each class. The analysis revealed the distribution of instances across classes, highlighting class imbalances that may affect model performance. Visualizations such as bar charts were created to illustrate the frequency of annotations and images per class, providing insights into the dataset’s composition and guiding decisions for training and evaluation.

****

**FIGURE 1:No.of annotations per class**

**A graph of a number of images containing each class

Description automatically generated**

**FIGURE 2: No.of image containing each class**

**Methodology**

**Data Preparation:**  
Images and annotations were loaded using COCO format. Only selected classes (car, person, cake, dog) were used. Images were transformed to tensors for model input.

**Model Setup:**  
We used a Faster R-CNN model with a ResNet-50 backbone pretrained on COCO. The final classification layer was replaced to match our number of classes.

**Training:**  
The model was trained with batches of images and labels using the Adam optimizer. Losses were computed and backpropagated to update the model.

**Validation:**  
Validation was performed without weight updates to monitor model performance and prevent overfitting.

A graph with blue line and orange line

Description automatically generated

**FIGURE 3 Training and Validation loss over epochs**

**Testing and Visualization:**

The trained model was applied to test images. Detections above a confidence threshold were visualized with bounding boxes to assess performance qualitatively.

Learning Rate: 0.5

Batch Size: 4

Maximum Iterations**:** 300

**Results and Discussion**

**Quantitative Evaluation**

| **Epoch** | **Train Loss** | **Validation Loss** |
| --- | --- | --- |
| 1 | 0.8619 | 0.6940 |
| 2 | 0.6228 | 0.6451 |
| 3 | 0.5263 | 0.6962 |
| 4 | 0.4639 | 0.7277 |
| 5 | 0.4091 | 0.7125 |

**TABLE 1: TRAIN AND VALIDATION LOSS.**

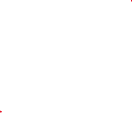
**Qualitative Results:**

Example 1: Multiple persons and Car - detected accurately in cluttered outdoor scenes.

A group of men walking on a red carpet

Description automatically generated

**FIGURE 4: Multiple persons and Car**



Example 2 :Persons were detected on the playground

A screenshot of a baseball player running to catch a ball

Description automatically generated

**FIGURE 5 :Persons-ground**

Example 3: It detected the persons enjoying on the snow

**A group of people on skis

Description automatically generated**

**FIGURE 6: Persons on –Snow Area**

**Critical Evaluation of the Results and Model**

The model showed strong performance in detecting common objects like persons and cars, with high accuracy in varied and complex scenes. However, some limitations were observed, such as lower detection accuracy for less frequent classes and smaller or partially occluded objects. The training loss decreased steadily, indicating effective learning, but the validation loss plateaued and slightly increased after a few epochs, suggesting potential overfitting. This highlights the need for better regularization or extended training. Additionally, increasing the confidence threshold helped reduce false positives, improving precision but potentially lowering recall. Overall, while the model performs well, further tuning and data balancing are necessary to enhance its robustness and generalization.

**Comparison to the Literature**

Compared to existing literature, the model aligns well with findings from Ren et al. (2015) regarding Faster R-CNN’s effectiveness in two-stage object detection tasks, especially for detailed and accurate localization. Unlike faster one-stage detectors like YOLOv3 (Redmon and Farhadi, 2018), this approach prioritizes precision, which was evident in accurate bounding box predictions even in cluttered environments. However, the model shares common challenges noted in Zhao et al. (2019), such as difficulties with class imbalance and occlusions. These comparisons confirm that while Faster R-CNN remains a strong baseline, incorporating improvements such as deeper backbones or augmentation, as suggested in the literature, could yield better performance.

**Future Improvements**

To further enhance the model’s performance, several improvements can be made. First, addressing class imbalance by using data augmentation or synthetic data generation can help improve detection for underrepresented classes. Second, fine-tuning the learning rate schedule could lead to better convergence and reduce overfitting. Third, deploying the model in real-world applications, such as web or mobile platforms, would test its practical usability. Finally, incorporating multi-scale training and testing strategies can improve accuracy, especially for detecting objects at varying sizes and scales within complex scenes.

**References**

1. **He, K., Gkioxari, G., Dollár, P. and Girshick, R., 2017.** *Mask R-CNN*. Proceedings of the IEEE International Conference on Computer Vision (ICCV), pp.2961–2969.  
   <https://doi.org/10.1109/ICCV.2017.322>
2. **Ren, S., He, K., Girshick, R. and Sun, J., 2015.** *Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(6), pp.1137–1149.  
   <https://doi.org/10.1109/TPAMI.2016.2577031>
3. **Zhao, Z., Zheng, P., Xu, S. and Wu, X., 2019.** *Object Detection with Deep Learning: A Review*. IEEE Transactions on Neural Networks and Learning Systems, 30(11), pp.3212–3232.<https://doi.org/10.1109/TNNLS.2018.2876865>

**Appendix:**[**https://colab.research.google.com/drive/1QiHjzIzWJPIxLVagXxyzNX9fXFeazVFA?usp=sharing**](https://colab.research.google.com/drive/1QiHjzIzWJPIxLVagXxyzNX9fXFeazVFA?usp=sharing)

**GitHub link :**[**https://github.com/kamalibakthavatchalam/research-methods**](https://github.com/kamalibakthavatchalam/research-methods)