Report on Object Detection Models for Pascal VOC and iMaterialist Datasets

Dataset Selection:

- **Datasets:**
  - **Pascal VOC 2007+2012**
  - **iMaterialist Fashion 2020 (from Kaggle)**

Data Subset Criteria and Distribution:

- **Criteria:**
  - For both datasets, the selection criteria focused on class balancing, ensuring each class was represented adequately to improve model performance and generalization. Each class within both datasets was sampled to include a consistent number of images per category.
- **Data Distribution:**
  - Below is a graph displaying the data distribution across selected classes, showing the balanced distribution of images per class, critical for fair model evaluation.

Model Selection and Reasoning:

- **Selected Models:**
  - **COCO-pretrained RetinaNet**: Chosen for its robust feature pyramid network, ideal for multi-scale object detection. This model was finetuned on both datasets to leverage the benefits of its anchor-based detection.
  - **COCO-pretrained R50-FPN Mask R-CNN**: This model was chosen due to its versatility in handling region proposals through feature pyramid networks (FPN) and its solid track record in object detection tasks. The Mask R-CNN architecture, although heavier, has shown to be effective in segmentation and detection tasks when finetuned.

2. Comparative Analysis

The two models' performances on both the Pascal VOC and iMaterialist datasets were evaluated based on key metrics: precision, recall, accuracy, and inference time.

- **Explanation of Results**:
  - Mask R-CNN performed better in both precision and recall across both datasets. Its improved performance can be attributed to its effective region proposal method, which captures object boundaries more accurately. The

inference time was slightly higher for Mask R-CNN, expected due to its more complex architecture compared to RetinaNet. RetinaNet, though faster, struggled slightly with complex object boundaries, leading to lower precision and recall scores.

3. Insights and Observations

- **Model Performance Observations:**
  - Mask R-CNN showed superior detection performance, especially on the Pascal VOC dataset, due to its region proposal mechanism that better handles varied object scales. RetinaNet was faster but tended to miss smaller objects in cluttered backgrounds, impacting its precision.
  - On the iMaterialist dataset, both models faced challenges due to diverse clothing styles and variations. However, Mask R-CNN still managed to retain higher performance, likely due to its segmentation approach, which helps in fine-tuning object boundaries.
- **Challenges and Solutions:**
  - **Challenge**: Data imbalance was initially an issue, especially in the iMaterialist dataset.
    - **Solution**: Balancing classes by downsampling overrepresented categories and augmenting underrepresented ones.
  - **Challenge**: Overfitting on smaller classes in the Pascal VOC dataset.
    - **Solution**: Early stopping during training and reduced learning rates were implemented to mitigate overfitting.

4. Screenshots of Front-End

Below are screenshots of the front-end interface.