**Title**

Comparative Analysis of State-of-the-Art Object Detection Models on Mainstream Datasets

**Introduction**

Object detection is a pivotal task in computer vision, having applications from many fields. The evolution of deep learning models has significantly advanced the accuracy and efficiency of object detection. This proposal outlines a project aimed at comparing the performance of several leading object detection models to identify the most suitable one(s) for application on the mainstream datasets, comprehensive datasets that includes variety of objects in diverse environments.

**Objectives**

* To evaluate and compare the performance of selected state-of-the-art object detection models on mainstream datasets(COCO, PASCAL VOC etc.).
* To analyze the models' efficiency in terms of detection speed and computational resource requirements.
* To provide insights into the strengths and limitations of each model, considering various factors such as object scale, occlusion, and environmental conditions.

**Methodology**

Models for Comparison

The project will focus on the following state-of-the-art object detection models:

* **YOLOv4 (You Only Look Once version 4):** Known for its speed and accuracy, YOLOv4 is ideal for real-time object detection tasks.
* **Faster R-CNN with ResNet-50 backbone:** A balance of speed and accuracy, this model uses a region proposal network, making it powerful for detecting small objects.
* **SSD (Single Shot MultiBox Detector) with MobileNetV3 backbone:** Optimized for mobile and low-power devices, offering a good compromise between speed and accuracy.
* **Mask R-CNN:** An extension of Faster R-CNN, adds a branch for predicting segmentation masks on each Region of Interest (RoI), useful for instance segmentation tasks.

Evaluation Metrics

The models will be evaluated using the following metrics:

* Precision and Recall
* Mean Average Precision (mAP) at different IoU (Intersection over Union) thresholds
* Inference time per image
* Model size and computational resource requirement

Experimental Setup

* All models will be trained on the same hardware configuration to ensure a fair comparison.
* Data augmentation techniques such as flipping, rotation, and scaling will be applied to enhance model robustness.

**Expected Outcomes**

* A comprehensive comparison of the selected models' performance on the mainstream datasets.
* Identification of the best-performing model(s) in terms of accuracy, speed, and efficiency.
* Insights into each model's suitability for specific types of object detection tasks or environments.

**Division of labor**

We have preliminarily decided that each person will be responsible for implementing and fine-tuning one specific detector and apply new features to it.

**Relationship to team members' backgrounds**

Jerry Wang: I’m a first year master student majoring in computer engineering with a focus on computer architecture. I have a previous internship experience with developing and deploying image segmentation and recognition deep learning models on embedded devices.

Hezi Jiang: I am a junior undergraduate student. I am working on an individual research project about motion prediction, and working with the Peoplewave team on basic data analysis.

Denny Li: I am a first year MCS student focusing on software development and game development . I have done several in class and individual projects on software and game development. This course is the first time I have ever interacted with machine learning because I want to at least have some understanding to the trend of the internet world. However, I do have some experience with the application of object detection in a IoT class but without the involvement of machine learning.

Heyang Li: I am a first-year MCS student with prior experience in the VRVC lab at ShanghaiTech University, working on CV/CG projects. I have contributed to work presented at conferences in SIGGRAPH and CVPR, focusing primarily on human/animal pose estimation.