Exploring Data Augmentation Techniques to Generate Synthetic Images for Improving Object Detection Performance

Kalpit Munot | Ruthvika Reddy Loka | Mahitha Pillodi |
Rishideep Reddy Rallabandi

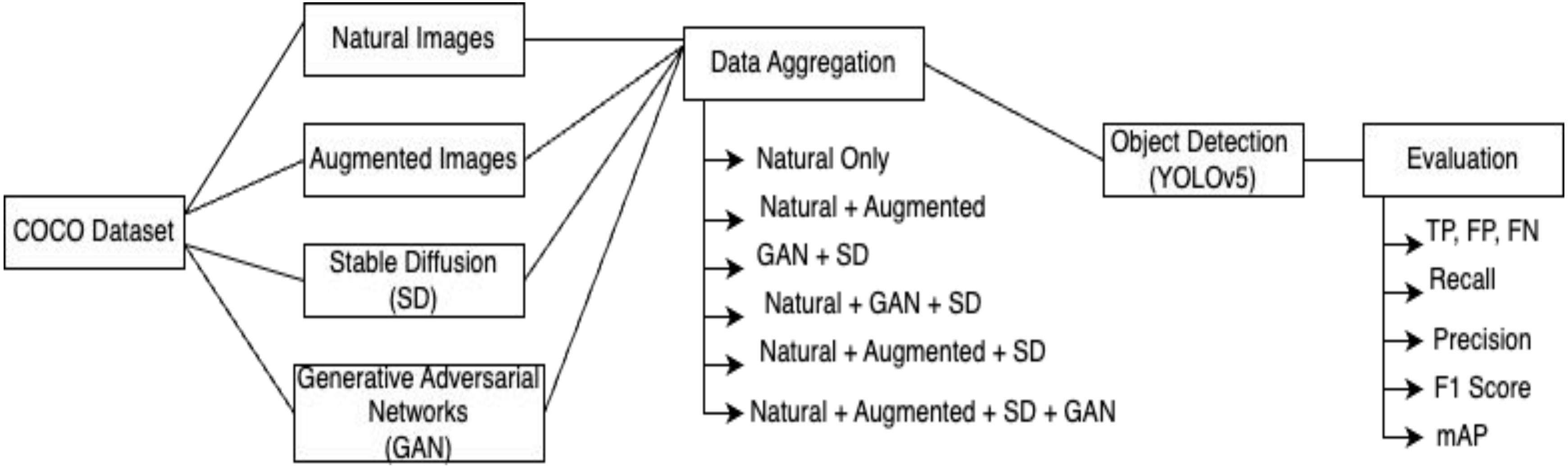
Problem Statement

- Object detection is critical in computer vision, but training deep learning models rely heavily on large and diverse datasets, which can be expensive and time-consuming to collect and annotate.
- To address this, data augmentation is a common technique. We propose a novel approach using Stable Diffusion (SD) and Generative Adversarial Networks (GANs) to generate synthetic images for augmentation.
- The objective is to assess the efficacy of the YOLOv5 algorithm and analyze how the model's performance is impacted by the integration of synthetic images.
- Our approach can be applied to various domains, including driving, surveillance, and medical imaging.

Current Work

- Current research on synthetic image generation is limited, with only a few studies exploring the use of synthetic images for object detection.
- Our work is novel in that it aims to utilize advanced techniques like Stable diffusion and GAN to generate realistic and diverse synthetic images for object detection.
- Our study fills a gap in the existing research by incorporating these techniques to generate new high-quality synthetic images and evaluating their effectiveness for object detection.

Experimental Design



Dataset Curation

- Our dataset consists of three parts: the original set, the synthetic set, and the test set. We used the COCO 2017 dataset for the original set, which contains over 118,000 images.
- We generated a synthetic set using Stable Diffusion and GANs and labeled them manually using makesense.ai tool.
- We also created a test set with images from a similar distribution to the original set.

Data Augmentation

Generative Adversarial Networks

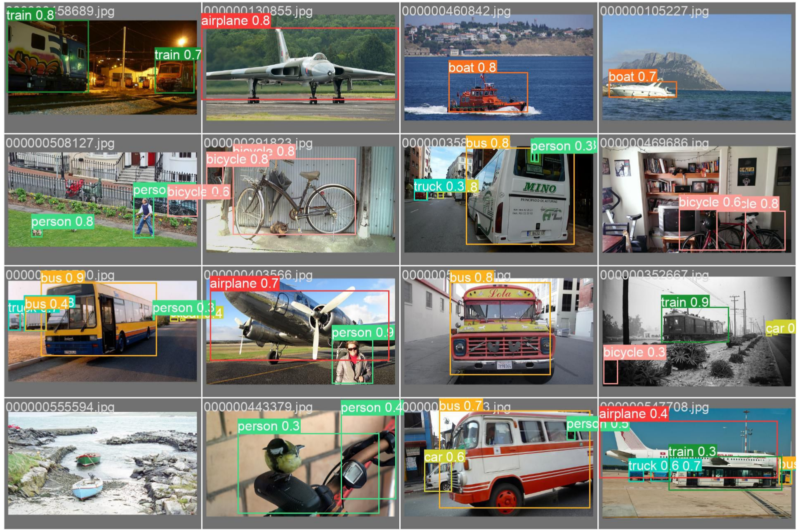
- GANs consist of a generator and a discriminator that are trained together to create synthetic data samples similar to real ones.
- We chose ICGAN model and fine-tuned it, which can produce targeted image outputs based on specific conditions. Our model can generate high-resolution and diverse images.
- We generated images by providing an input image for reference or specifying an input image class for reference.
- We created images with different contexts and scenarios for training.

Stable Diffusion




- Stable diffusion image creation is a process that generates visually appealing and coherent images using diffusion models.
- We used stable-diffusion-v1-4 pre-trained model from Hugging face.
- A neural captioning model was used to generate neural captions which was fed to this model to generate images

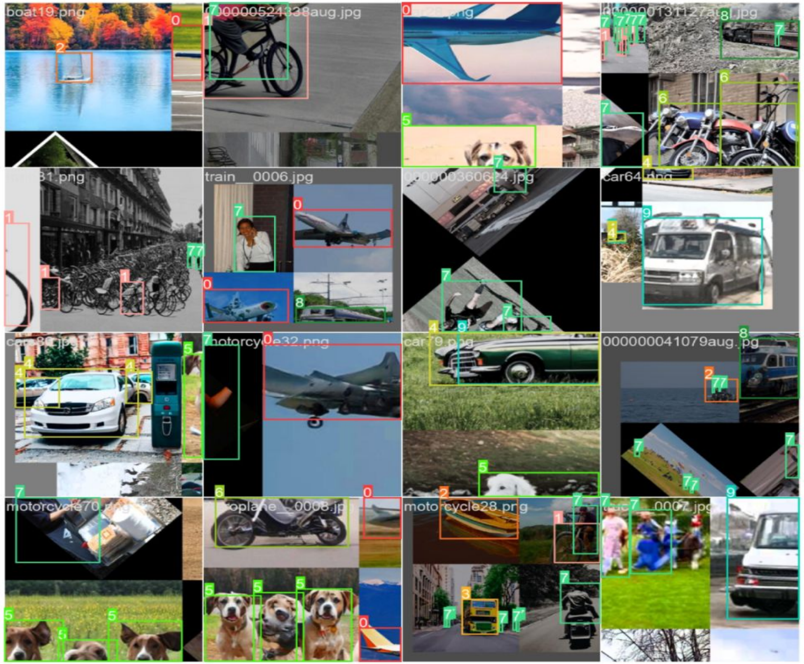
Object Detection - YOLOv5



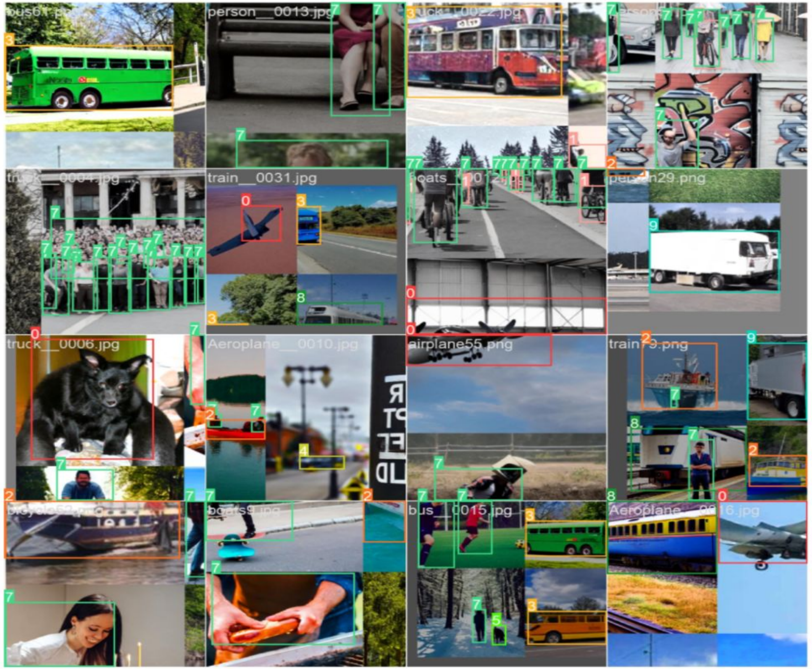
- We used the YOLOv5 algorithm for real-time object detection and applied transfer learning to train a custom dataset. We combined original images from the COCO dataset with synthetic images generated from a pre-trained stable diffusion model to improve diversity and model performance.
- The model was trained and results were evaluated using manually generated annotations.
- Our experimental design demonstrates nearly same accuracy and efficiency like natural with synthetic and augmented datasets as well.

Results

Object Detection Results



NATURAL+GAN+SD+AUG

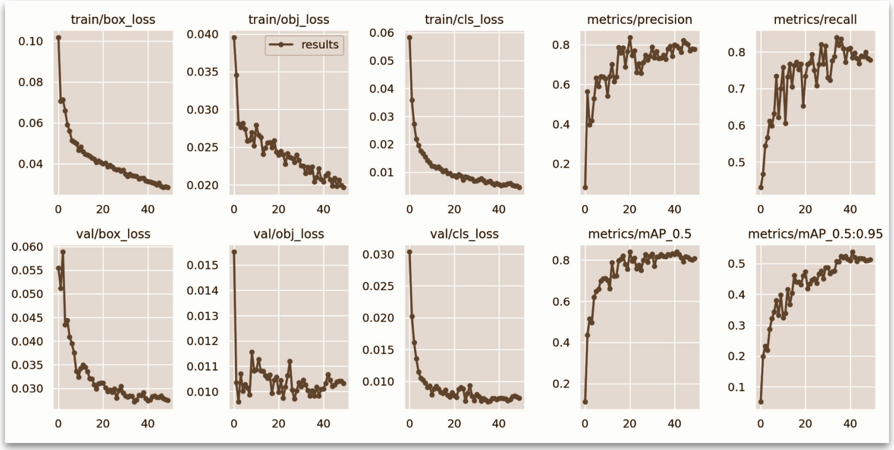


GAN+SD

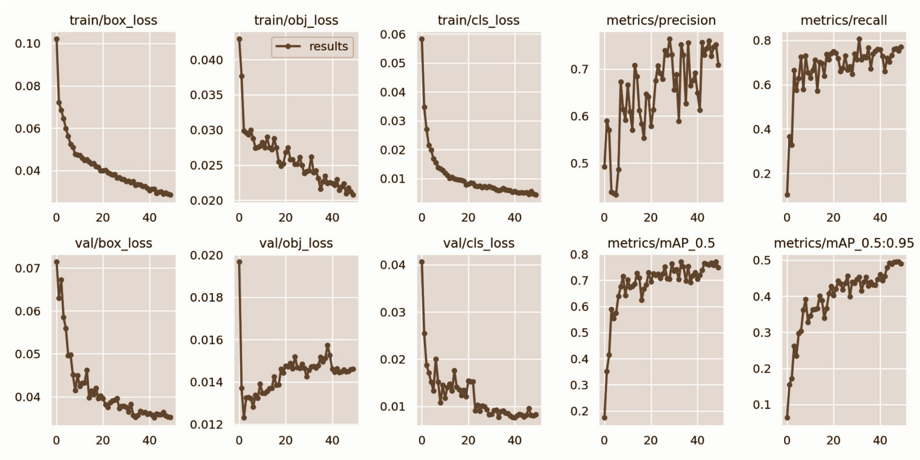
Evaluation Metrics

- TP,FP,FN :- True Positives, False Positive, False Negative
- Recall:- TP / (TP + FN)
- Precision:- TP / (TP + FP)
- F1 Score:- 2 * (Precision * Recall) / (Precision + Recall)
- mAP:- $(AP_1 + AP_2 + ... + AP_n) / n$

Results

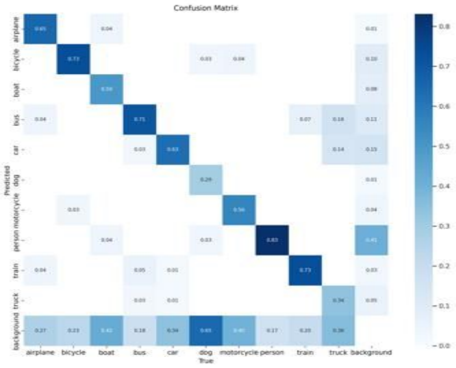


NATURAL + GAN + SD + AUG

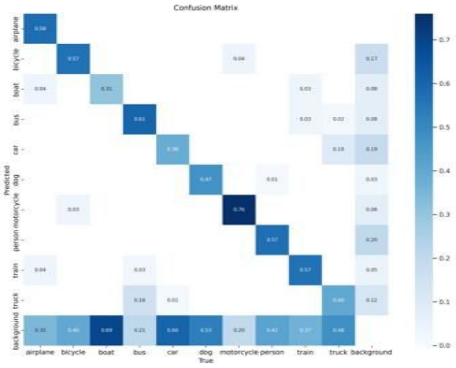


NATURAL + SD + AUG

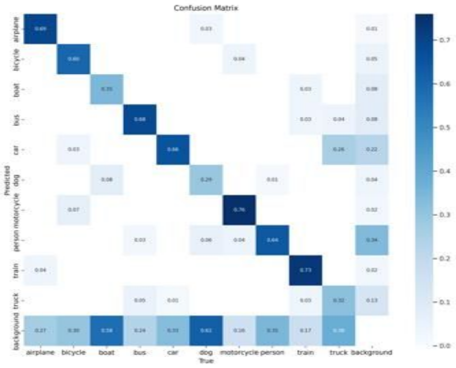
Confusion Matrix



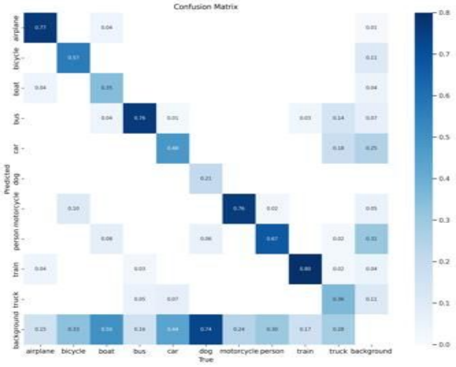
NATURAL



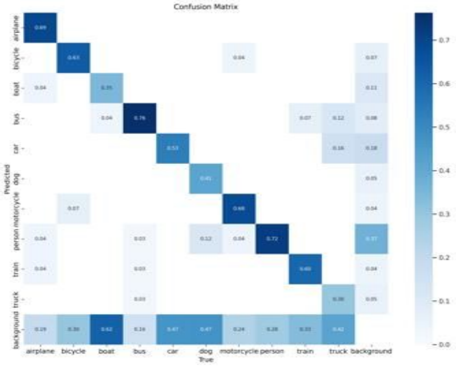
GAN + SD



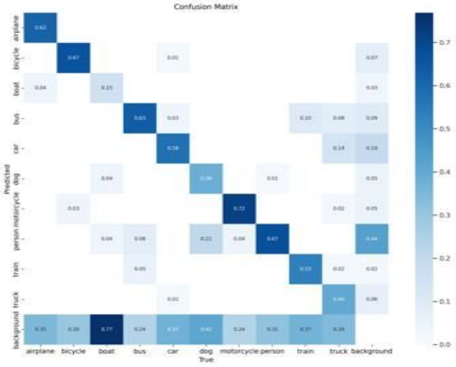
NATURAL + GAN + SD



NATURAL + AUG

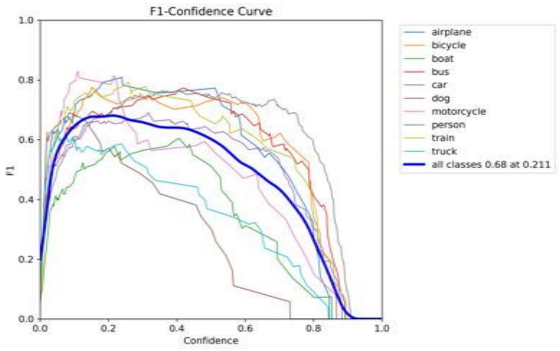


NATURAL + SD + AUG

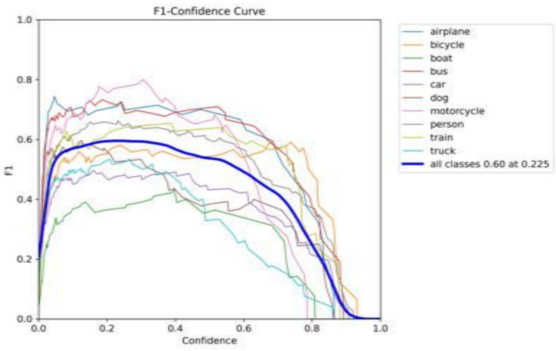


NATURAL + GAN + SD + AUG

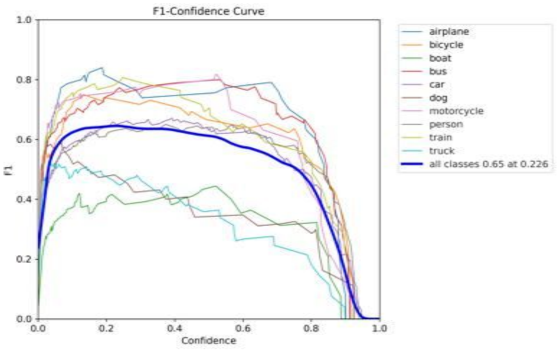
F1-Confidence Curve



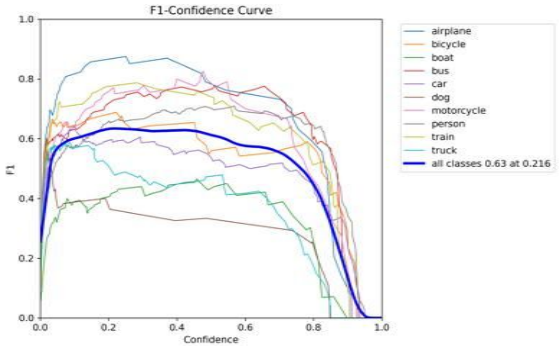
NATURAL



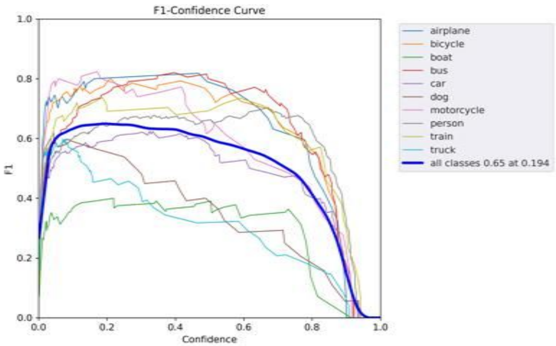
GAN + SD



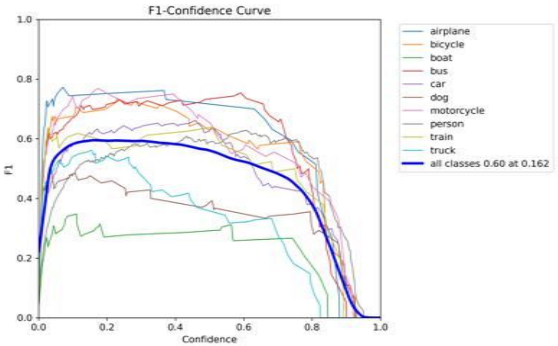
NATURAL + GAN + SD



NATURAL + AUG

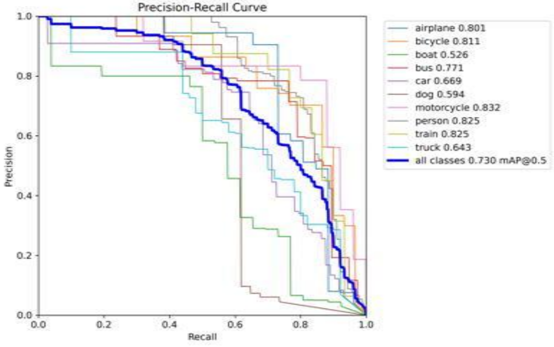


NATURAL + SD + AUG

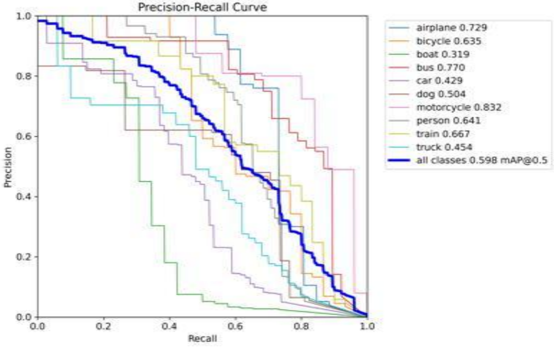


NATURAL + GAN + SD + AUG

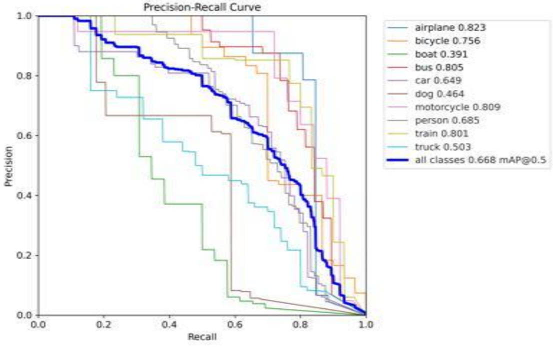
PR Curve



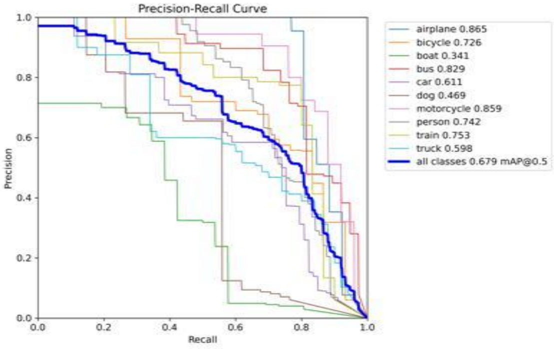
NATURAL



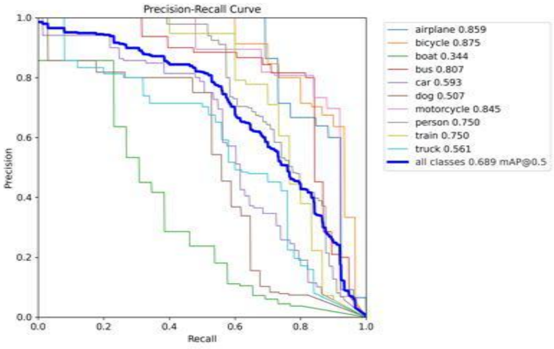
GAN + SD



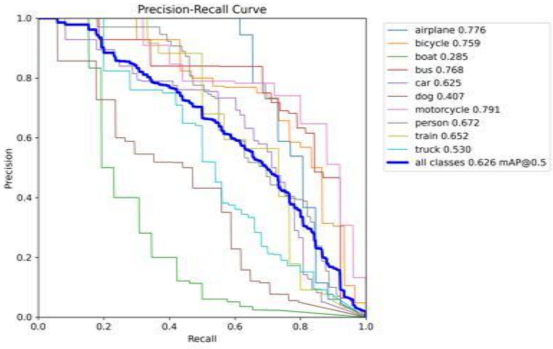
NATURAL + GAN + SD



NATURAL + AUG

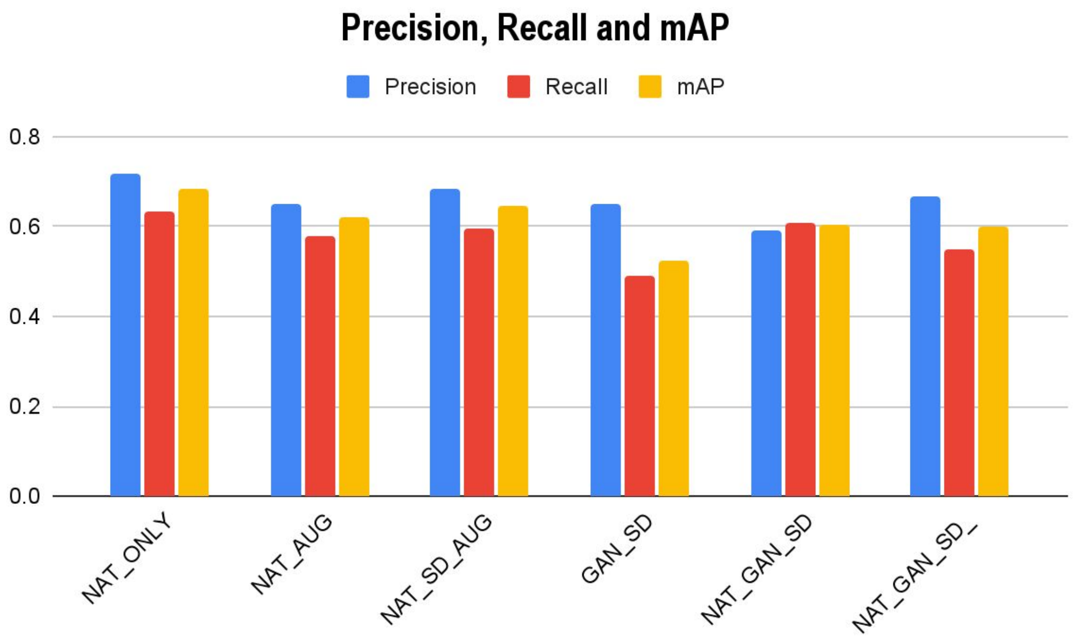


NATURAL + SD + AUG



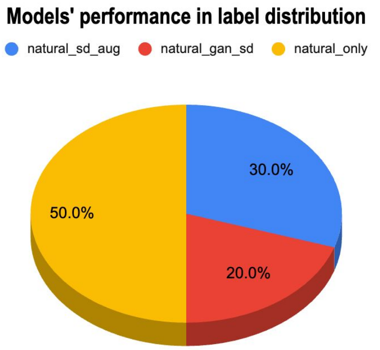
NATURAL + GAN + SD + AUG

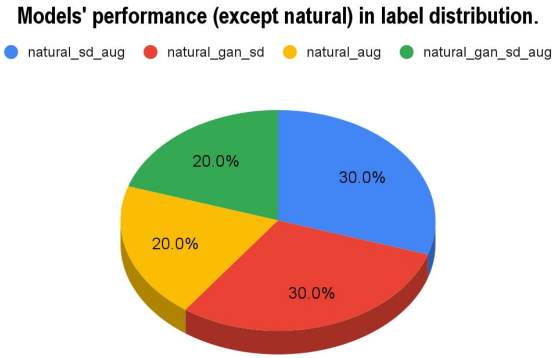
mAP Graph

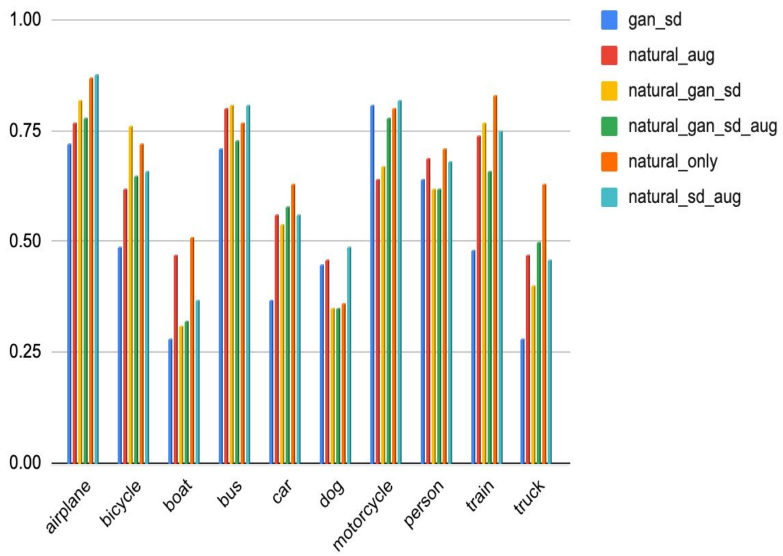


Precision, Recall and mAP
Precision
Recall
mAP
0.8
0.6
0.4
0.2
0.0
NAT_ONLY
NAT_AUG
NAT_SD_AUG
GAN_SD
NAT_GAN_SD
NAT_GAN_SD_

Model's performance on different datasets







Conclusion

- Synthetic images yield comparable results to natural images for object detection tasks, making it a valuable technique.
- GAN technique exhibits lower performance than other synthetic datasets, possibly due to distorted images.
- Results can be replicated with advanced techniques like Stable Diffusion and GANs by reducing the dataset by 40-60%.
- We gained knowledge about different data augmentation methods and their implementation.
- We acquired insights into techniques for enhancing object detection performance.

Future Work

- We could explore the potential of using other advanced generative models, such as VAEs and flow-based models, to generate synthetic images for object detection.
- The study could be extended to evaluate the effectiveness of the proposed approach on different object detection algorithms, such as Faster R-CNN and RetinaNet.
- The impact of different augmentation parameters, such as rotation and scaling, could be analyzed to determine their influence on the model's performance.
- The approach could be applied to different domains, such as robotics and autonomous vehicles, to assess its effectiveness in diverse real-world scenarios.


Thank
You