Exploring the Effectiveness of Dataset Synthesis

Alexander van Meekeren
Supervisor: Maya Aghaei Gavari

Summer 2023

Introduction

Traditional data acquisition for training deep learning models can be time-consuming. Our proposed solution is synthetic data generation through image synthesis, specifically focusing on apple orchard images. By utilizing Stable Diffusion models [1], we aim to create images for training deep learning models. This research aims to enhance data collection processes and improve efficiency in various industries. Resulting in the following research questions:

- How might latent diffusion models, such as Stable Diffusion, be utilized to create apple orchard images?
- What is the performance of a deep learning model trained on synthetic data versus a deep learning model trained on real data?

Abstract

Deep object detection models have made significant strides, but the need for large training datasets remains a challenge. To address this, we explore the feasibility of using synthetic data generated by Stable Diffusion for apple tree object detection. Comparing it to a baseline model trained on real data, we create hyper realistic apple tree datasets using prompt engineering and pretrained Stable Diffusion. The annotated dataset is then used to train a YOLOv5m model [3]. Results show that the generated data model performs similarly to the baseline on real-world images, with a small average precision difference. Synthetic data generation offers potential as an alternative for training deep object detection models, reducing the reliance on extensive real data collection.

Figure 1. Image of a real-world apple tree from the MinneApple dataset [2]

Figure 3. Samples of the custom synthesized dataset of apple trees, using Stable Diffusion 2.1-base.

Experiments and Results

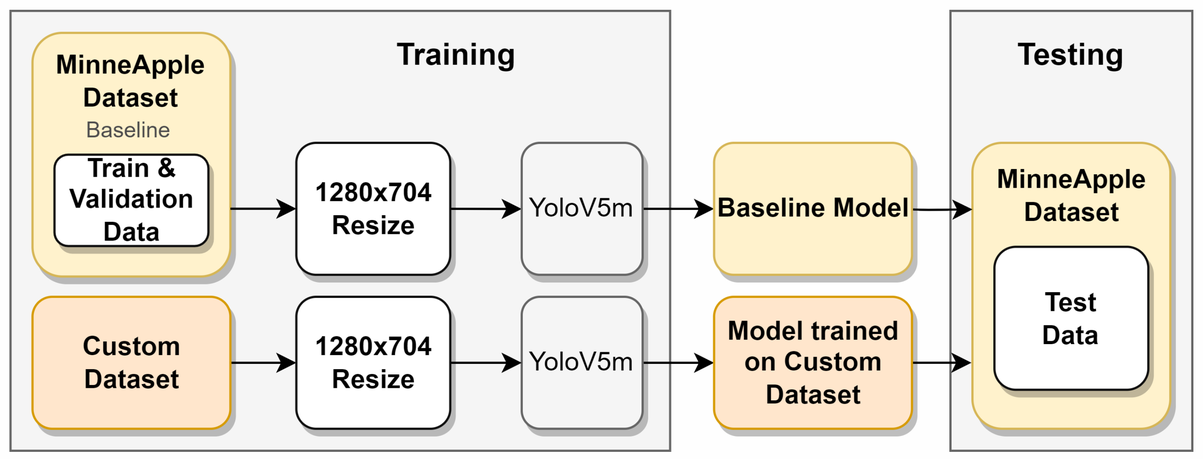

Figure 4. Overview of the experiments preformed in the research.

- A baseline model underwent training using real-world data.
- Another model was trained on the custom-synthesized dataset.
- The performance of both models was assessed on real-world data, specifically the MinneApple test dataset, using the Average Precision evaluation metric.

Table 1. An overview of the average precision obtained by testing the trained models (both baseline and generated) on the MinneApple test dataset, along with their corresponding differences.

Dataset	AP@0.50	AP@0.5:0.05:0.95	AP@0.75
Baseline	0.70	0.36	0.34
Generated	0.61	0.30	0.25
Difference	***0.09***	***0.06***	***0.09***

Materials and Methods

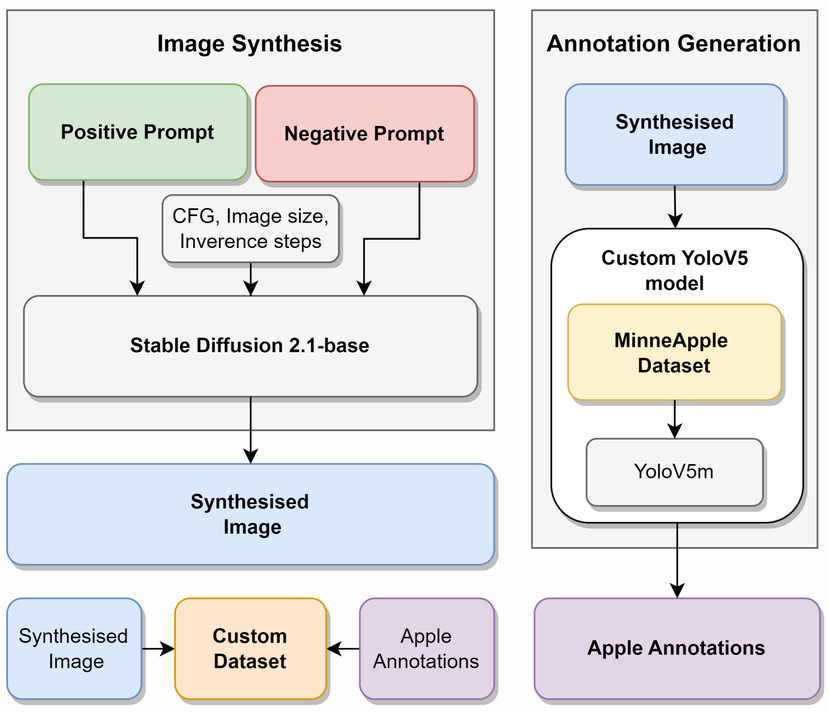

Figure 2. illustrates the process of creating a synthesized **custom dataset**.

Positive Prompt:
"a photo of a tree standing in the grass. the tree has many apples, the apples are both red and yellow. beneath the tree there are a lot of apples. The many apples are a combination of red apples and yellow apples. volumetric lighting. shadows, hyper realism, 4k, realism, photograph“

Negative Prompt:
"blurry image, deformed, cartoon, drawing, painting"

- The **image synthesis** stage involved generating custom images using the Stable Diffusion model, utilizing Positive and Negative prompts as well as Hyperparameters.
- In the **Annotation Generation** stage, the generated images were annotated using a YoloV5m model trained on the existing apple tree dataset called MinneApple.
- The synthesized images along with their annotations were combined to produce a unique **custom dataset**.

Conclusion

The study proposal has revealed the promise of employing artificially generated images to improve the current data collection process. This is achieved by comparing a model trained on synthesized data with a model trained on real-world data.

- Deep learning models, specifically YOLOv5m in this research, trained on synthetic data, demonstrate a comparable level of performance to models t rained on real-world data.
- Stable Diffusion can synthesize apple tree images to training deep learning algorithms, by adjusting given prompts and parameters like CFG and image size.
- Continuing to explore Stable Diffusion models for image synthesis can advance data collection and deep learning training, opening new opportunities and accelerating progress in multiple fields.

NHL STENDEN
computer vision & data science

Acknowledgements

- This project was financially supported by Regieorgaan SIA (part of NWO) and performed within the RAAK PRO project Mars4Earth.
- We would like to thank our collaborators at Saxion University of Applied Sciences for insightful discussions

Regieorgaan SIA
SAXION UNIVERSITY OF APPLIED SCIENCES

References

- [1] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Bj¨orn Ommer. High-resolution image synthesis with latent diffusion models, 2021.
- [2] Nicolai Hani, Pravakar Roy, and Volkan Isler. MinneApple: A benchmark dataset for apple detection and segmentation. IEEE Robotics and Automation Letters, 5(2):852–858, apr 2020.
- [3] Glenn Jocher et al. ultralytics/yolov5: v6.2 - yolov5 classification models, apple m1, reproducibility, clearml and deci.ai integrations.