

EXAMENSARBETE Evaluating synthetic data for enhancement of object detection models**STUDENT** Carl Wikström**HANDLEDARE** Volker Krueger (LTH), Patric Fröjd (Verisure Innovation)**EXAMINATOR** Jacek Malec (LTH)

Exploring the Power of Synthetic Data in Detecting Humans

POPULÄRVETENSKAPLIG SAMMANFATTNING **Carl Wikström**

Synthetic data is being explored as a powerful tool to enhance object detection models, offering a promising alternative to traditional real-world datasets. But how powerful is synthetic data when training an object detection model for more specific tasks?

To delve deeper into the capabilities of synthetic data in object detection, we decided to examine its impact on the YOLOv9c model, a widely-used tool known for its efficiency in detecting objects within images. We began by training the model on synthetic data, which allowed us to generate extensive and diverse training scenarios not readily available in real-world datasets. Following this, we fine-tuned the model on specific real-world datasets to evaluate how well it could adapt and perform in practical situations. This two-step process aimed to explore the potential of synthetic data to improve detection accuracy and robustness in real-world applications.



Figure 1: Synthetic data images

Our study specifically targeted human detection in security surveillance contexts, a domain where

real-world data can be both limited and challenging to acquire. By introducing variations in the synthetic data—such as changes in human placements, poses, and quantities—we sought to determine how these factors influence model performance. This approach provided insights into how well synthetic data can prepare a model for the complexities of real-world scenarios and enhance its overall effectiveness.

In our evaluation, we compared the performance of models trained solely on synthetic data with those trained exclusively on real data. The goal was to understand the benefits of using synthetic data as a preliminary training resource and its impact on model accuracy when subsequently fine-tuned with real-world data.

Our results revealed that the performance of synthetic data varies significantly depending on its composition, highlighting the importance of selecting the right data mix. Notably, synthetic data can match real data performance when the model is fine-tuned with a small portion of the target dataset. Additionally, synthetic data proved valuable in improving detection capabilities for previously unseen scenarios, such as crawling people, demonstrating its ability to enhance model robustness in diverse situations.