Cyber-Physical Systems Final Project: Anomaly Detection in Autonomous Vehicle Camera Data Using Variational Autoencoders

Mark Herrmann, Mason Hawkins, Johntae Leary

Spring 2025

Abstract

This report explores anomaly detection in autonomous vehicle (AV) camera sensor data using a Variational Autoencoder (VAE). The project utilizes the Audi Autonomous Driving Dataset to train a VAE model on normal camera images and detect anomalies that could indicate faulty or malicious data. We outline the overlap of computing, networking, and physical processes in AV, explain the problem of sensor data anomalies, and discuss its importance for safety and security. The solution employs a three-layer CPS architecture, detailing the physical, network, and application layers, alongside data collection and processing methods. We address security challenges such as data injection and spoofing, and explore their potential harm, which can cause system failures or safety hazards. The VAE-based solution demonstrates promising results, correctly identifying anomalies in simulated noisy data, enhancing AV reliability and cybersecurity.

1 Introduction to Cyber-Physical Systems in the Project

1.1 Type of CPS

This project focuses on Autonomous Vehicle Cyber-Physical Systems, specifically targeting the camera sensor data from the front-center camera of an autonomous vehicle. AV CPS integrate advanced computing, networking, and physical processes to enable vehicles to perceive their environment, make decisions, and execute actions without human intervention. The camera sensor, a critical component, captures visual data essential for object detection, lane keeping, and navigation.

1.2 Integration of Computing, Networking, and Physical Processes

In AV CPS, computing involves onboard processing units (i.e., GPUs, TPUs) that run artificial intelligence models, such as the Variational Autoencoder used in

this project. These AI models process sensor data and detect anomalies. Networking within AVs facilitate internal communication between sensors, actuators, and computing units through protocols like Controller Area Network bus, and external communication with cloud servers using TCP/IP or 5G networks. Physical processes include the vehicle's sensors, including cameras, LIDAR, and radar. Actuators or components that produce movement from a signal can be seen in the form of steering, braking systems, and acceleration. The VAE processes camera images to ensure data integrity, directly impacting the vehicle's perception and decision-making, thus integrating these CPS components to enhance safety and reliability.

2 Problem Statement

2.1 Specific Problem

The project addresses the problem of detecting anomalies in AV camera sensor data, specifically images from the front-center camera. Anomalies, or in this case distorted images due to hardware faults or malicious data injection, can compromise the vehicle's perception and accuracy, leading to wrong decisions and potential accidents.

2.2 Relevance

Anomalies in camera sensor data are a critical issue in AV CPS because they threaten the vehicle's ability to accurately interpret its environment. For instance, a distorted image might cause the vehicle to misidentify obstacles or lane markings, leading to collisions. This problem is relevant as AVs rely heavily on sensor data for real-time decision-making, and any deviation from normal data patterns can have severe consequences, including loss of life, increased distrust in AV technology, and legal ramifications.

3 Importance of the Problem

3.1 Criticality

Addressing anomalies in AV sensor data is vital for the CPS industry in order to maintain the safety, security, and reliability of these vehicles. AVs are already revolutionizing transportation, but their adoption hinges on public trust in their ability to operate safely. Undetected anomalies can lead to faulty decisions and catastrophic outcomes, such as collisions or pedestrian injuries.

3.2 Potential Benefits

Solving this problem enhances:

- **Safety**: Detecting anomalies prevents accidents caused by faulty or malicious data, ensuring safer roads.
- **Security**: Identifying compromised data protects against cyberattacks, such as data injection, bolstering AV cybersecurity.
- **Reliability**: Consistent anomaly detection ensures robust performance, increasing user confidence in the vehicle.
- **Efficiency**: Early detection reduces downtime and maintenance costs by flagging issues before they escalate.

The VAE-based anomaly detection system acts as an early warning mechanism, safeguarding AV operations and fostering trust in autonomous technologies.

4 Solution Overview

4.1 Three-Layer Architecture Approach

4.1.1 Physical Layer

The physical layer comprises hardware components, including the front-center camera, which captures RGB images of the vehicle's environment. Other sensors, such as LIDAR and radar, complement the camera but are not the focus of this project. Actuators, including steering motors and braking systems, execute commands based on processed sensor data. The camera's role is to provide visual input for perception tasks, which the VAE analyzes for anomalies to ensure data integrity before the image reaches components that control movement.

4.1.2 Network Layer

The network layer handles communication between the camera, onboard computing units, and other vehicle systems. Internally, the camera data is transmitted via the CAN bus, a robust protocol for low-latency communication in AVs. Externally, the vehicle may communicate with cloud servers or Vehicle to-everything systems using TCP/IP over 5G for updates and collaborative decision-making. The VAE operates on the onboard computing unit, processing camera data locally to minimize latency and ensure real-time anomaly detection.

4.1.3 Application Layer

The application layer includes the VAE model, implemented with TensorFlow/Keras, which processes camera images to detect anomalies. The software interfaces with the vehicle's perception system to provide alerts when anomalies are detected. The VAE utilizes data from the physical layer in the form of images and transmitted data in the network layer to deliver functionality, such as flagging distorted images. Similar approaches using VAE-based architectures have been applied in control systems for signal interpretation and decision-making [1]. This

layer also supports potential user interfaces, such as dashboards displaying anomaly alerts for human operators or developers.

4.2 Data Collection and Processing

4.2.1 Data Collection

Data is collected from the A2D2 dataset, specifically the preview dataset with 120 camera-lidar images for training and 10 camera-lidar-semantic-bboxes subset with 10 images for validation. The camera-lidar images are unnanotated and the camera-lidar-semantic-bboxes images have semantic segmentation and bounding boxes. The front-center camera captures RGB images, which are preprocessed by undistorting (correcting for lens distortion) and resizing to 256x160 pixels to make the data more lightweight and manageable. The training data represents ”normal” scenes, while validation data is used after training the model to create test anomalies by adding noise.

4.2.2 Data Processing

Data processing involves:

- **Preprocessing**: Images are converted to RGB, undistorted using OpenCV, resized, and normalized to [0, 1].
- **VAE Processing**: The VAE encodes images into a latent space, reconstructs them, and calculates reconstruction errors. High errors indicate anomalies.
- **Real-Time Requirements**: The VAE is optimized for onboard processing, and would be pretrained when implemented in an AV.

The VAE learns normal image patterns during training, enabling it to detect deviations in test data. Anomalies are simulated by adding Gaussian noise with a factor of 0.4 to validation images, mimicking faulty or attacked data.

5 Security Challenges

5.1 Security Issues

Two critical security issues in AV CPS are:

1. **Data Injection**: Attackers may inject fake objects such as a stop sign/pedestrian into camera feeds, misleading the vehicle’s perception system.
2. **Sensor Spoofing**: Malicious actors could manipulate camera inputs to provide false environmental data, such as altered lane markings.

5.2 Criticality and Stakeholder Trust

These issues are critical because they exploit vulnerabilities in the perception system, undermining the vehicle's ability to make safe decisions. Data injection can cause the vehicle to brake unnecessarily or ignore real obstacles, while spoofing can lead to navigational errors. Such attacks erode overall trust in EVs, including public confidence in AV safety and manufacturer credibility, and could lead to large profit loss, potentially halting AV adoption.

5.3 Consequences of Failure

Failing to address these issues could result in:

- **System Failure**: Incorrect perception leads to faulty decisions, such as collisions or dangerous maneuvers.
- **Safety Hazards**: Accidents caused by anomalies could injure passengers, pedestrians, or other road users.
- **Legal/Financial Repercussions**: Manufacturers could face lawsuits, recalls, and damage to reputation.
- **Worst-Case Scenario**: Attackers could remotely induce accidents, causing loss of life and catastrophic financial losses for manufacturers and owners of AVs.

The VAE mitigates these risks by detecting anomalies in real time, flagging suspicious inputs before they affect decision-making.

6 Results and Discussion

The VAE was trained on 120 A2D2 images and validated on 10 images, achieving a validation loss that indicates effective learning of normal patterns. After 27 Epochs the VAE was able to achieve a reconstruction loss(MSE) of below 1000 on the training data and a reconstruction error on the validation set of 4300. This indicates the model was overfitting the training data, however, I was unable to achieve better results with my current computing power. The VAE took 13 minutes to train over 27 minutes. In testing, it correctly detected 5/5 simulated anomalies (noisy images) with 1/5 false positives, demonstrating decent accuracy. However, in a real-world situation, there would need to be a much higher accuracy due to the high-risk nature of driving. This could be easily accomplished with greater computing power and a larger dataset used for training. The reconstruction error threshold (95th percentile of normal errors) effectively separated normal and anomalous data, as visualized in error distribution histograms. These results suggest the VAE is a viable tool for real-time anomaly detection, enhancing AV CPS security. Similar hybrid approaches combining VAEs and GANs have been shown to improve anomaly detection performance in other sensor-based systems [2].

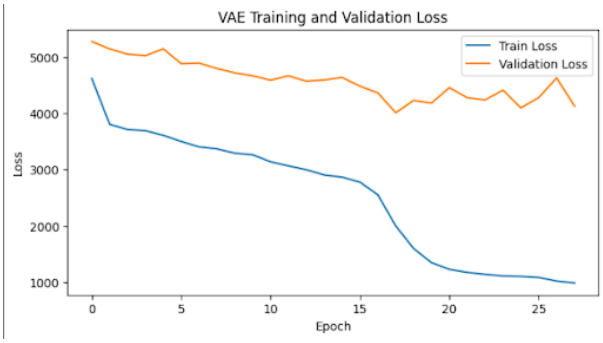

Figure 1: Training and Validation Error over 27 Epochs

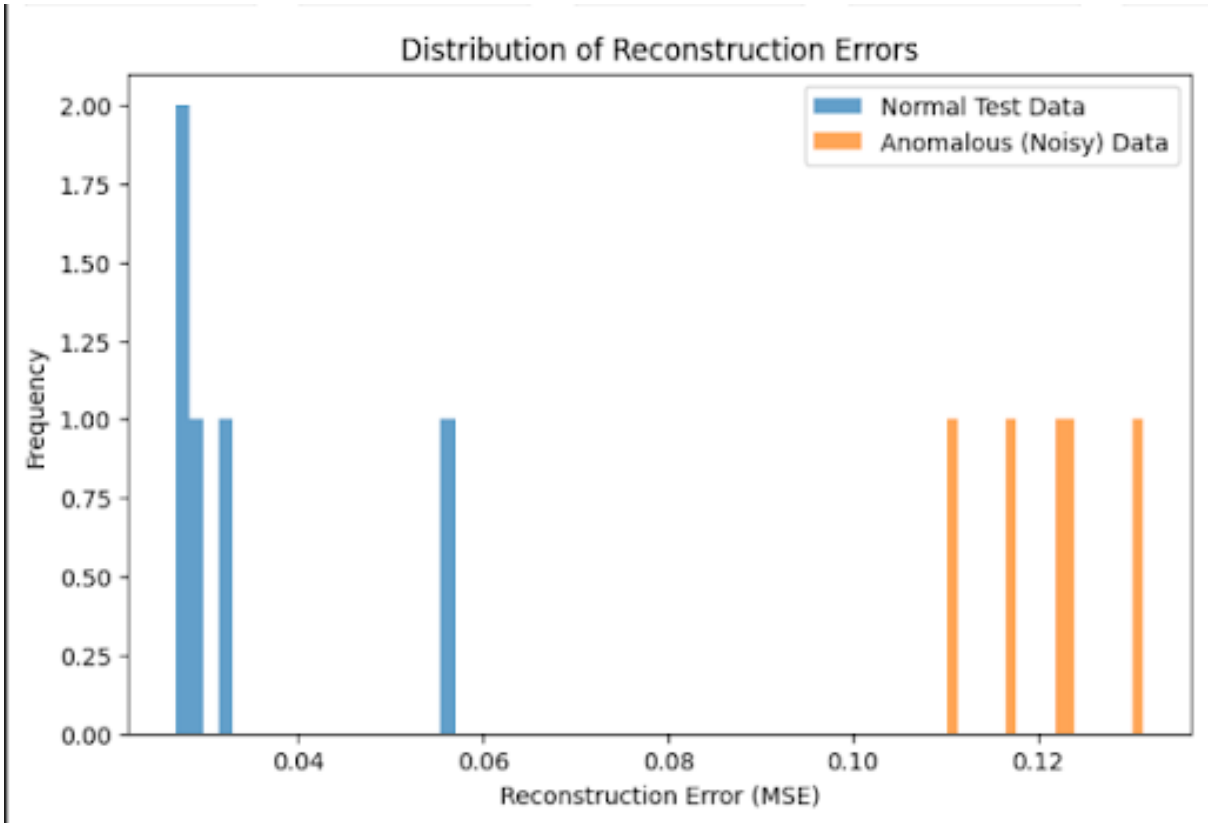

Figure 2: Reconstruction Error of Anomalies and Normal images

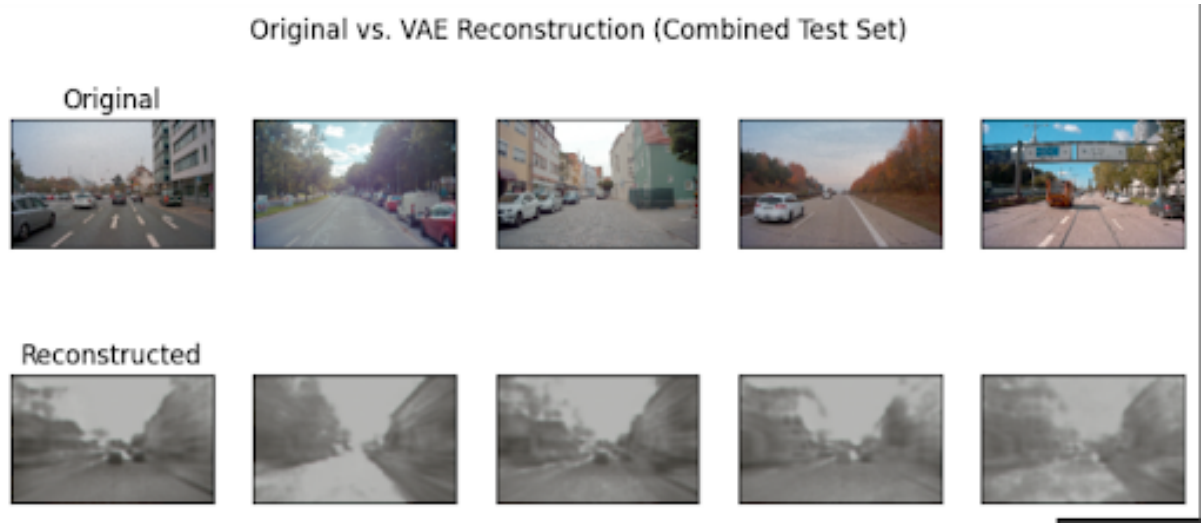

Figure 3: Reconstructed images after VAE

7 Conclusion

This project demonstrates the efficacy of a VAE for anomaly detection in AV camera data, addressing a critical CPS challenge. By integrating computing, networking, and physical processes, our solution enhances safety, security, and reliability in AVs. The three-layer architecture and robust data processing ensure real-time performance, while the VAE's ability to flag anomalies mitigates security risks like data injection and spoofing. Future work could explore training the model on multiple sensor's data, more complex anomaly detection(object input like stop signs), and automated responses to further strengthen AV cybersecurity. Anomaly detection can pave the way for safer autonomous transportation.

References

[1] Keith A. Currier. Variational autoencoder and sensor fusion for robust myoelectric controls, 2023. Accessed: May 3, 2025.

[2] Xiao Wang and Han Liu. Data supplement for a soft sensor using a new generative model based on a variational autoencoder and wasserstein gan. *Journal of Process Control*, 85:91–99, 2020.