Real-Time Highway Traffic Analysis and Generative Adversarial Network Based Anomaly Detection

M.Tech Thesis

by

Aditya Chandle

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

INDIAN INSTITUTE OF TECHNOLOGY INDORE

June 2025

Real-Time Highway Traffic Analysis and Generative Adversarial Network Based Anomaly Detection

A THESIS

Submitted in partial fulfillment of the

requirements for the award of the degree

of

Master of Technology

by

Aditya Chandle

2302101004

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

INDIAN INSTITUTE OF TECHNOLOGY INDORE

June 2025

INDIAN INSTITUTE OF TECHNOLOGY INDORE

CANDIDATE'S DECLARATION

I hereby certify that the work which is being presented in the thesis entitled **Real-Time Highway Traffic Analysis and Generative Adversarial Network Based Anomaly Detection** in the partial fulfillment of the requirements for the award of the degree of **Master of Technology** and submitted in the **Department of Computer Science and Engineering, Indian Institute of Technology Indore,** is an authentic record of my own work carried out during the period from July 2023 to May 2025 under the supervision of Prof. Aruna Tiwari, Indian Institute of Technology Indore, India, Dr. Sanjay Singh, CSIR-CEERI,Pilani,India and Dr. Sumeet Saurav, CSIR-CEERI,Pilani,India
The matter presented in this thesis has not been submitted by me for the award of any other degree of this or any other institute.

Aditya

7/June/2025
Signature of the Student with Date

(Aditya Chandle)

This is to certify that the above statement made by the candidate is correct to the best of my knowledge.

9/6/2025
Signature of Thesis Supervisor # 1 with Date

(Prof. Aruna Tiwari)

07/06/2025
Signature of Thesis Supervisor # 2 with Date

(Dr. Sanjay Singh)

Sumeet Saurav 07/06/25
Signature of Thesis Supervisor # 3 with Date

(Dr. Sumeet Saurav)

Aditya Chandle has successfully given his M.Tech. Oral Examination held on **30/April/2025**.

07/06/2025
Signature(s) of Supervisor(s) of M.Tech. thesis Date:

Subhra Mazumdar
Signature of Chairman, PG Oral Board with Date

Date: 09.06.2025

09/06/25
Signature of HoD with Date

Date:

ACKNOWLEDGEMENTS

I would like to take this opportunity to express my gratitude to people who in one or the other way contributed by this academic journey time as meaningful, enriching, and fulfilling. At first, I would like to thank my supervisors **Prof. Aruna Tiwari**, **Dr. Sanjay Singh** and **Dr. Sumeet Saurav** who were a constant source of inspiration during my work. Without their constant guidance and research directions, this research work could not be completed. Their continuous support and encouragement have motivated me to remain streamlined in my research work.

I would also like to express my sincere thanks to **Dr. Rituraj Singh** for his constant support, valuable insights, and encouragement throughout the project. His guidance has been instrumental in shaping the direction and outcomes of my research.

I am thankful to all the **faculty members and staff of the Department of Computer Science and Engineering, IIT Indore** for providing a conducive environment for research and academic growth. Their assistance and cooperation have always been appreciated.

I would like to express my heartfelt respect and gratitude to my **family and friends** for their unconditional love, care, and support. Their patience, understanding, and encouragement have given me strength and motivation during every phase of this journey.

Finally, I am thankful to all who directly or indirectly contributed, helped, and supported me in completing this work.

Dedicated to My Family

ABSTRACT

This thesis presents the design and development of an advanced video surveillance system aimed at enhancing highway traffic monitoring and anomaly detection. The proposed system integrates multiple traffic monitoring capabilities including multi-axle vehicle classification, speed estimation, and vehicle counting, which are essential for understanding and managing vehicular activity on highways. These features are individually developed using deep learning and computer vision methodologies to ensure accurate and real-time performance even in challenging environments.

The system is further extended by incorporating an anomaly detection module based on Constrained video anomaly detection GAN (CVAD-GAN), which leverages latent space modeling to learn the normal patterns in traffic video frames. The CVAD-GAN framework reconstructs video frames and uses pixel-wise reconstruction error in conjunction with discriminator feedback to detect deviations from the learned normal behavior, thereby identifying anomalies in real-time video streams. During testing, reconstruction is used to localize abnormal regions in video frames, making the system robust and practical for deployment.

This thesis also discusses the complete training pipeline of the CVAD-GAN architecture, explaining its behavior as a autoencoder and the mathematical formulations that govern its optimization. The integration of this anomaly detection mechanism with standard traffic analytics provides a comprehensive solution for intelligent surveillance systems.

Overall, the developed system is suitable for highway traffic authorities, toll plazas, and smart city applications where automatic and accurate traffic monitoring, along with real-time anomaly detection, is required. Hence , it is applied to real toll plaza , problem of National Highways Authority of India(NHAI) through Devaditya Technocrats pvt. Ltd. ,Indore. This work contributes to the field of intelligent transportation systems by providing a modular, scalable, and deep learning–driven framework for video surveillance in dynamic environments.

Contents

List of Figures

List of Tables

List of Abbreviations

CVAD-GAN	Constrained Video Anomaly Detection via Generative Adversarial Network
GAN	Generative Adversarial Network
VAD	Video Anomaly Detection
ML	Machine Learning
DL	Deep Learning
CNN	Convolutional Neural Network
TP	True Positive
FP	False Positive
FN	False Negative
TPR	True Positive Rate
FPR	False Positive Rate
ROC	Receiver Operating Characteristic
AUC	Area Under the Curve
YOLO	You Only Look Once
GPU	Graphics Processing Unit
CPU	Central Processing Unit
VAE	Variational Autoencoder

Chapter 1

Introduction

The explosive growth of video streams from surveillance cameras has given rise to an imperative demand for smart systems that can automatically interpret and analyze scenes. Human monitoring is not only labor-intensive but also subject to human error, especially in high-density or high-speed settings like highways or city intersections. Consequently, computer vision-based automatic video analysis through deep learning has become a viable contender, with encouraging applications in the detection of events such as traffic offenses, vehicle classification, and anomalies [13, 9]. Here, the integration of object detection, behavior modeling, and anomaly detection models has become imperative in order to increase the efficiency and reliability of contemporary surveillance systems.

1.1 Background

Anomaly detection in video monitoring involves the detection of patterns or activities that greatly differ from regular action. Conventional machine learning approaches have proved to be less scalable and generalizable in actual real-world complicated environments [20, 11]. Deep learning-based methods, including Convolutional Neural Networks (CNNs), provide improved feature representation and have proved effective in large-scale video classification and abnormal event detection[9, 21].

Generative models, especially Generative Adversarial Networks (GANs), have proven to be quite successful in capturing normal patterns of behavior and detecting deviations[3, 15, 7] . The advent of models such as Spatio-Temporal Adversarial Networks (STAN) and Latent Space Autoregression has also enhanced temporal perception in videos[10, 1]. At the same time, advances such as attention mechanisms and geometric transformations have amplified anomaly localization abilities[22, 4].

In the context of intelligent traffic surveillance, deep learning techniques are also playing a crucial role in real-time vehicle tracking, speed estimation, and multi-axle vehicle classification. Vehicle tracking ensures continuous monitoring of motion trajectories, enabling behavioral analysis across frames. Speed detection leverages temporal information between consecutive frames to estimate the velocity of moving vehicles, which is vital for identifying overspeeding or unusual driving patterns. Meanwhile, multi-axle classification distinguishes between light and heavy vehicles based on structural features, aiding in traffic flow analysis, toll management, and infrastructure planning. These advancements not only complement anomaly detection systems but also broaden their applicability in smart highway monitoring and traffic regulation. In this thesis, all these requirements are identified from an industrial collaboration Devaditya Technocrats pvt. Ltd. Indore as they are handling real problem assigned by NHAI(National Highway Authority Of India).

1.2 Motivation

The motivation behind this work stems from the need to address specific challenges in video surveillance on highways, including poor lighting, occlusions, motion blur, and top-down camera angles. These issues hinder accurate detection and classification of vehicles, especially when distinguishing between similar classes such as multi-axle and standard trucks. Moreover, the detection of subtle anomalies like sudden stops, reverse movements, or slowdowns often goes unnoticed in real-time systems.

Recent advancements in deep anomaly detection models[20, 21, 7] and attention-based frameworks[22, 12] offer promising tools to build robust, real-time surveillance solutions.

However, there remains a significant gap in integrating these models into an end-to-end highway monitoring system that can simultaneously detect, track, classify, and identify abnormal patterns effectively. Currently, commercially available intelligent cameras lack the advanced capabilities required for comprehensive traffic monitoring and analysis. This highlights the need for a robust, intelligent solution tailored specifically for highway traffic surveillance and analytics.

1.3 Objectives

The main objectives of this thesis are:

- To develop an automated and modular highway surveillance system with deep learning methodologies.
- YOLO-based object detection for real-time vehicle recognition and tracking.
- To employ axle-based classification for vehicle type differentiation under top-view observation.
- To propose video-based speed estimation techniques with robustness against environmental factors.
- To combine GAN-based and attention-guided anomaly detection models for detecting anomalous traffic behaviors.
- To make use of LoRA for minimizing computational overheads
- To deploy the proposed system on GPU-accelerated edge computing platforms.

1.4 Thesis Contribution

This thesis contributes the following:

- A real-time vehicle detection and classification pipeline optimized for top-view highway video.

- A new axle-based classification approach resistant to occlusions and visual noise.
- A parallelized framework for speed estimation and multi-camera video processing.
- The integration of GAN architectures for spatio-temporal anomaly detection.
- The use of LoRA to minimize computational overhead in model adaptation without compromising accuracy.
- The complete system with multi-axle classification, speed detection, vehicle counting, and anomaly detection deployed on GPU-accelerated hardware for real-time performance.

1.5 Organization of the Thesis

- Chapter 1 gives the introduction, background, motivation, and contributions of the thesis.
- Chapter 2 gives an extensive literature review covering object detection, anomaly detection, and deep learning frameworks applied in surveillance systems, with a particular emphasis on models designed to process and analyze video data effectively.
- Chapter 3 details the dataset acquisition, preprocessing pipeline, and annotation strategies essential for training deep learning models in vehicle detection, axle classification, and anomaly detection using both benchmark and real-world surveillance data.
- Chapter 4 discusses the evaluation metrics, implementation details, and experimental results for anomaly detection using the CVAD-GAN model and its efficacy across benchmark datasets and real-world highway applications.
- Chapter 5 details the design and implementation of an integrated traffic monitoring system featuring modular components for vehicle detection, multi-axle classification, speed estimation, real-time visualization, and the integration of an anomaly detection module (CVAD-GAN) with the Advanced Traffic Counter and Classifier (ATCC) for

enhanced traffic event monitoring.

- Chapter 6 highlights the system's limitations and outlines possible directions for future research and deployment improvements.

Chapter 2

Literature Review

This chapter supplies an extensive overview of the major concepts, models, and preprocessing methods involved in the suggested anomaly detection framework. Section 2.1 describes the video dataset preprocessing pipeline, such as frame extraction, resizing, normalization, and frame selection strategies required for acquiring high-quality training and testing sets. Section 2.2 supplies the basic principles behind Generative Adversarial Networks (GANs), particularly CVAD-GAN in Section 2.2.1. It also addresses the performance metrics utilized in testing model performance under Section 2.2.2, and it points out the specific application of CVAD-GAN to highway traffic monitoring under Section 2.2.3. Section 2.3 looks into the application of Low-Rank Adapters (LoRA) for cost-effective fine-tuning in GAN-based models, which improves adaptability with low computational expense. Lastly, Section 2.4 gives an introduction to YOLOv8, the real-time object detector model utilized for car tracking and axle categorization in the proposed system.

2.1 Video DataSets Preprocessing

To train and test the performance of the designed object detection and anomaly detection modules, both benchmark and real-world datasets are employed. Anomaly detection components are trained over public surveillance datasets such as UCSD Pedestrian (Peds1 and Peds2), CUHK Avenue, ShanghaiTech, and Subway datasets. The datasets involve labeled normal and anomalous events and offer a credible benchmark to measure the generalization of GAN-based models [2, 14, 16].

Furthermore, self-recorded surveillance videos retrieved from toll plazas on Indian highways are utilized to fine-tune the YOLOv8 model. The videos contain heterogeneous traffic conditions with miscellaneous lighting, climatic conditions, and types of vehicles, such as heavy multi-axle trucks. The fine-tuning improves the detection capability of YOLOv8 in the exact working environment of the suggested system.

The data preprocessing activities involved:

- Frame Extraction: Fixed sampling rate keyframes were extracted to produce training and testing samples
- Resolution Standardization: The frames were rescaled to 640×640 pixels to match the input YOLOv8 requirements.
- Label Formatting: The annotations were reformatted into YOLO format with normalized coordinates.
- Augmentation: Random flipping, brightness scaling, and Gaussian noise addition were used to enhance generalization.
- Motion Feature Extraction: Optical flow and frame differencing were employed to inform training of the GANs by emphasizing motion patterns important for anomaly detection [5].

These preprocessing operations ensured the resilience and flexibility of the models on both benchmark and real-world data sets.

2.2 Generative Adverserial Network

Generative Adversarial Networks (GANs), proposed by Ian Goodfellow et al. in 2014, are a type of deep learning network that is specifically suited for generative tasks, where the aim is to learn the data distribution to generate realistic synthetic examples. A GAN is formed of two neural networks, known as the generator and the discriminator, which are trained competitively in parallel. The generator tries to create the data close to the actual data, and the discriminator attempts to discriminate between real and generated sample. This adversarial procedure leads the generator to continually enhance its capacity to create high-fidelity data. GANs have achieved notable success in a wide range of applications such as image generation, video synthesis, super-resolution, and anomaly detection because they can model complicated data distributions without defining them explicitly[6].

2.2.1 CVAD-GAN: Constrained Video Anomaly Detection via Generative Adversarial Network

In the context of highway surveillance, where the monitoring of complex vehicle behavior is essential, CVAD-GAN [17]presents a highly effective solution for anomaly detection. Designed as a lightweight architecture, CVAD-GAN[17] consists of two primary components: a generator and a discriminator. The generator is further divided into an encoder-decoder structure. During processing, video frames augmented with white Gaussian noise are passed through the encoder, which extracts latent features. These features are constrained using Kullback–Leibler (KL) divergence to ensure they conform to a predefined distribution. The decoder then reconstructs the video frame from these constrained latent features.

The training of CVAD-GAN[17] is carried out exclusively on normal video frames—those devoid of any anomalous events. Consequently, when the model encounters an anomalous input during inference, the generator, having never seen such deviations, fails to reconstruct the abnormal regions accurately. This results in a reconstructed frame that lacks the anomalous components present in the original frame. The discrepancy between the original and reconstructed frames is thus used to localize and identify anomalies within the scene.

This reconstruction-based anomaly detection approach is particularly suitable for highway surveillance environments, where detecting unexpected behaviors such as wrong-way driving, illegal halts, or sudden lane changes is critical.

2.2.1.1 Methodology

In the context of highway surveillance, the proposed **CVAD-GAN**[17] model processes an input video frame X by adding white Gaussian noise, resulting in a modified input $\tilde{X}$, which is then used to reconstruct a corresponding frame $\hat{X}$. The model incorporates two types of noise for distinct purposes: (1) the noise η, added to the input frame, improves the generalization capability of the generator, and (2) the noise N is used to constrain the latent space during training. The generator G is responsible for learning a meaningful representation of normal traffic frames, while the discriminator D is trained to differentiate between frames drawn from the true data distribution (p_{normal}) and those produced by the generator (p_t).

Throughout adversarial training, the generator attempts to produce reconstructions that the discriminator cannot distinguish from real inputs, thereby minimizing the difference between X and $\hat{Y}$. However, since the generator is trained solely on normal traffic data, it fails to accurately reconstruct anomalous events—such as wrong-way driving or vehicles stopped in motion lanes—resulting in noticeable deviations in $\hat{X}$. These distortions serve as a reliable indicator for identifying abnormal behavior in surveillance footage, making **CVAD-GAN**[17] an effective component of the proposed anomaly detection system.

2.2.1.2 Training Pipeline of CVAD-GAN

The complete training pipeline of the CVAD-GAN[17] model is illustrated in Figure 2.1, which outlines the generator and discriminator interaction for detecting anomalies in video frames.

1. The input video frame $Y \sim p_{\text{normal}}$ is passed through the encoder G_1 of the generator

after adding white Gaussian noise η:

$$G_1(Y+\eta) \rightarrow Z \sim \mathcal{N}(0,\sigma)$$

This step improves generalization by denoising.

2. The latent representation Z is constrained using Kullback-Leibler (KL) divergence to match a standard normal distribution:

$$\text{KL}(Z \parallel \mathcal{N}(0,1))$$

The decoder G_2 then reconstructs the frame:

$$G_2(Z \sim \mathcal{N}(0,\sigma)) \rightarrow \hat{Y} \sim p_{\text{normal}}$$

3. For all $Y \sim p_{\text{normal}} + \eta$ and $Y \nsim p_{\text{normal}} + \eta$, the following condition holds:

$$P(G_1(Y)|N) \geq P(G_1(\hat{Y})|N)$$

4. The generator is optimized to minimize the reconstruction loss $\mathcal{L}_G$, resulting in:

$$G(Y+\eta) \rightarrow \hat{Y} \sim p_{\text{normal}}$$

Thus, it behaves like a denoising autoencoder.

5. For an anomalous frame $\tilde{Y} \nsim p_{\text{normal}}$, the generator maps it to an undefined distribution:

$$G(\tilde{Y}+\eta) \rightarrow \hat{Y} \sim p_{\text{undefined}}$$

Consequently, $\mathcal{L}_G$ does not converge to zero and the anomaly is highlighted by the distortion in the output.

6. The discriminator D is trained to distinguish between real and generated frames:

$$\mathcal{L}_D = -\left[\log D(Y) + \log(1 - D(\hat{Y}))\right]$$

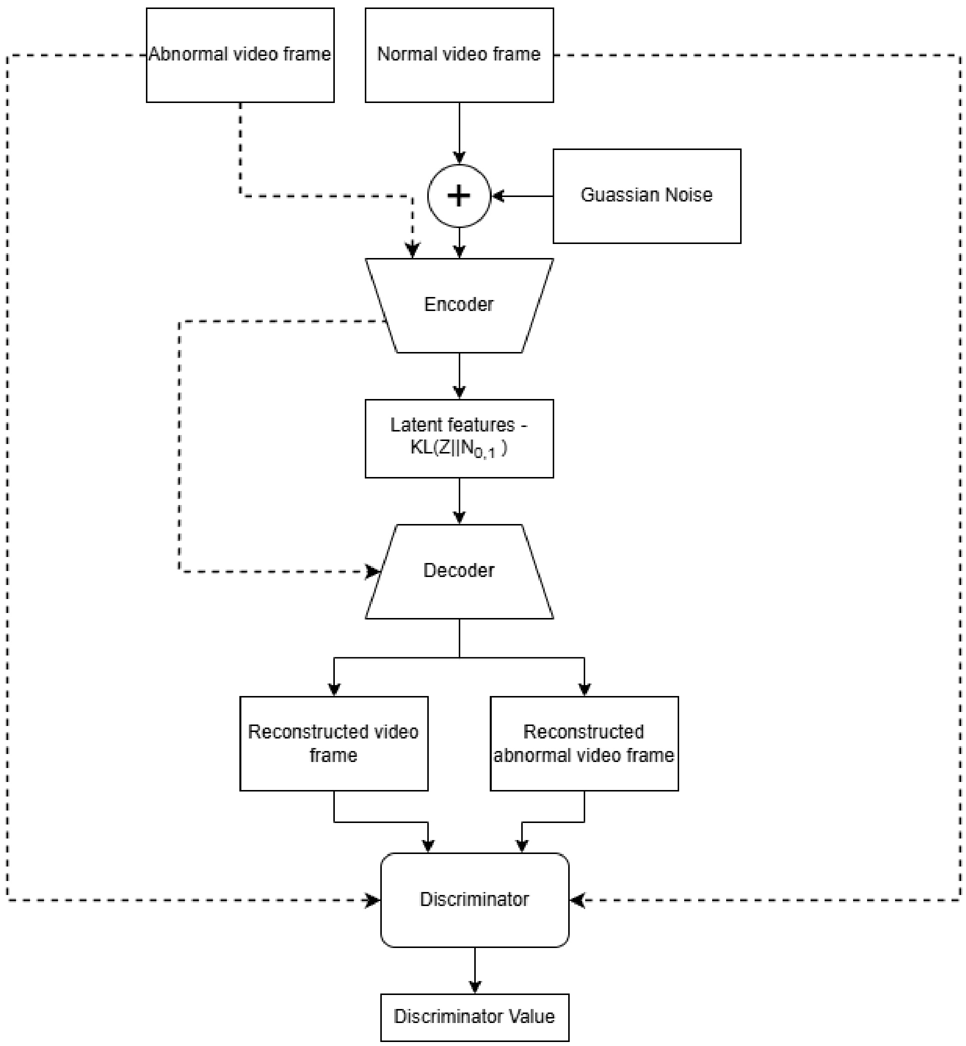


Figure 2.1: Proposed Training Pipeline of CVAD GAN.

2.2.2 Evaluation Metrics

Following the evaluation protocols adopted in prior works [17], we assess anomaly detection performance at the frame level. The primary evaluation tool is the Receiver Operating Characteristic (ROC) curve, which measures the discriminative capacity of the model. The ROC curve plots the True Positive Rate (TPR) against the False Positive Rate (FPR), calculated as follows:

$$\text{TPR} = \frac{TP}{TP + FN} \tag{2.1}$$

$$\text{FPR} = \frac{FP}{FP + TN} \tag{2.2}$$

Here, TP, FN, FP, and TN are the number of true positives, false negatives, false positives, and true negatives, respectively. Besides the Receiver Operating Characteristic (ROC) curve, the Equal Error Rate (EER) and the Area Under the ROC Curve (AUC) are given to give a complete assessment of model performance. In addition, the Area Under the Precision-Recall Curve (AU-PR) is calculated, which is especially handy for measuring performance on imbalanced data sets. The AU-PR is calculated as:

$$\text{AU-PR} = \int_0^1 \text{Precision}(R)\, dR \tag{2.3}$$

where R denotes the recall, and precision is the proportion of true positives among all predicted positives. A higher AU-PR implies stronger anomaly detection capabilities.

2.2.3 Application of CVAD-GAN in Highway Traffic Video Anomaly Detection

In the context of highway traffic monitoring, **CVAD-GAN**[17] (Constrained video anomaly detection via generative adversarial network) offers a powerful solution for detecting anomalous events in surveillance videos, such as sudden halts, wrong-way driving, over-speeding, or vehicles stopped at undesignated locations. This model learns the spatio-temporal patterns of normal highway traffic by training exclusively on video sequences that depict regular vehicular motion and flow. The generator, guided by a conditional variational autoencoder, synthesizes future frames based on historical inputs, capturing both spatial structure and temporal continuity. Simultaneously, the adversarial discriminator evaluates the authenticity of these predicted frames, forcing the generator to produce outputs that closely match real, normal traffic behavior.

At the inference stage, when an input video contains events that deviate from the learned distribution—such as a vehicle entering from the wrong lane or stopping abruptly—the generator fails to reconstruct these anomalies accurately. This results in high reconstruction errors and increased discriminator loss, which can be used to flag these events as abnormal. Such a framework is especially valuable on highways, where early detection of atypical vehicle behavior is critical to prevent accidents, manage traffic flow, and ensure the safety of commuters. By focusing on unsupervised learning from normal data, CVAD-GAN eliminates the need for exhaustive anomaly annotations, making it scalable and adaptive for real-world highway surveillance scenarios.

2.3 Low-Rank Adapter (LoRA) for Efficient Fine-Tuning in GAN-Based Models

Based on the basic architecture of Generative Adversarial Networks (GANs) explained in Section 2.2, this section presents **Low-Rank Adaptation (LoRA)**[8] as an efficient parameter-adaptive mechanism to optimize the adaptability and scalability of GAN-based models like CVAD-GAN [17]and VALD-GAN[18].

LoRA[8] facilitates efficient fine-tuning by introducing two latent low-rank trainable matrices $A \in \mathbb{R}^{r \times d}$ and $B \in \mathbb{R}^{d \times r}$ into the update of latent weights, so that the update of weights becomes:

$$\Delta W = BA \quad \text{where } r \ll d$$

This enables the model to learn new tasks without having to retrain the entire network. Notably, the pre-trained model weights are not updated, and only the adapter layers are trainable through backpropagation. This cuts the number of trainable parameters way down—by more than 90

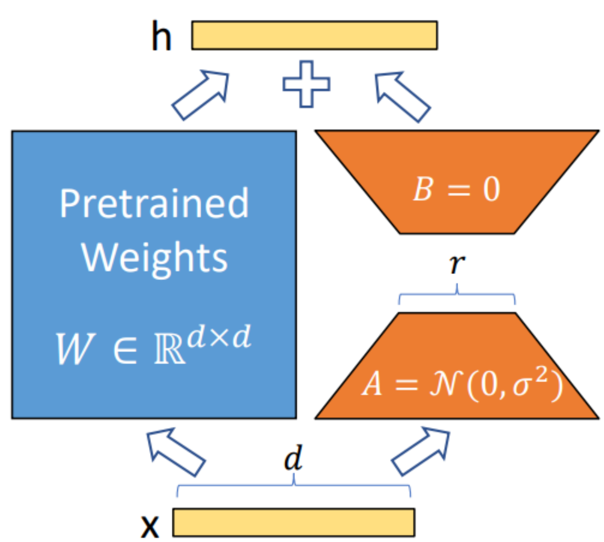


Figure 2.2: Low-Rank Decomposition: The weight matrix W is reparemterized in terms of the product of two low-rank matrices A and B. Only A and B are trained whereas W is kept frozen[8].

In video anomaly detection with GANs, LoRA is particularly useful. It enables the generator and discriminator networks to be optimally fine-tuned for domain-specific issues—like changing illumination, vehicle models, and environmental conditions of highway surveillance systems—without heavy computational costs.

This light and modular variant is especially well-suited for deployment in edge computing scenarios such as toll booths and traffic stops, where real-time performance and resource scarcity are paramount. Integrating LoRA into the GAN architecture presented in Section 2.2, the system proposed is able to maintain adaptability as well as efficiency in high-resolution video surveillance tasks.

2.4 YOLO V8

YOLOv8 (You Only Look Once, version 8) is the newest and most sophisticated version in the YOLO series, created by Ultralytics. It is based on the fundamental design concepts of the earlier versions but adds a few enhancements, including an anchor-free detection mechanism, decoupled detection heads, and a simplified backbone for increased inference speed and accuracy. YOLOv8 employs a CSPDarknet backbone coupled with PANet for efficient feature aggregation, and is capable of real-time performance even on edge devices.

In highway surveillance, YOLOv8 is utilized for object detection and multi-object tracking. Its coupling with Deep SORT enables the vehicle trajectories to be tracked consistently between frames, rendering it very applicable for traffic flow analysis and monitoring of vehicle behavior [19]. The capability of the model to deal with object scales and occlusions also improves its reliability in densely populated or high-speed scenes prevalent in highway videos.

Additionally, YOLOv8 is extremely flexible and allows for custom training with little computational cost. Here, it is fine-tuned on a domain-specific dataset of highway videos to enhance detection accuracy for classes like multi-axle trucks and two-wheelers. Its real-time performance, along with its robustness to environmental changes, makes it an essential part in the proposed automated traffic monitoring system.

Chapter 3

Dataset Acquisition and Preprocessing

This chapter is structured into four key sections that collectively detail the data preparation process for training and evaluating the proposed object detection and anomaly detection modules. Section 3.1 (Dataset Description) introduces the datasets used in this study, encompassing both publicly available surveillance benchmarks—such as UCSD Pedestrian (Peds1 and Peds2), CUHK Avenue, ShanghaiTech, and Subway—and real life high-resolution 4K CCTV footage collected from highway toll plazas by Devaditya Technologies Pvt. Ltd Indore. This section emphasizes the significance of using both controlled and real-world data to ensure the models' generalizability. Section 3.2 (Preprocessing Pipeline) elaborates on the automated workflow developed to manage the computational challenges of processing large video files. It covers key preprocessing steps including frame extraction at fixed intervals, resolution resizing to 640×640 pixels for YOLOv8 and 160×160 grayscale images for CVAD-GAN, label formatting into YOLO-compatible annotations, and the application of data augmentation techniques such as horizontal flipping, brightness scaling, and Gaussian noise addition. Additionally, this section discusses motion feature extraction through optical flow and frame differencing, which is particularly important for training GAN-based models on spatio-temporal anomalies. Section 3.3 (Data Annotation) describes the manual labeling of approximately 2,500 frames using annotation tools to define bounding boxes for five key classes: wheel, truck, motorcycle, car, and bus. These annotations serve as ground truth for both axle classification and vehicle detection tasks.

An illustrative example of an annotated frame is also included to demonstrate the labeling structure. Finally, Section 3.4 (Summary) recaps the chapter by highlighting the importance of a robust data pipeline, which combines benchmark datasets with practical toll plaza footage, thorough preprocessing, and accurate annotations to enable effective training of deep learning models in varied traffic scenarios.

3.1 Dataset Sources

3.1.1 Benchmark Datasets for Anomaly Detection

A few popular benchmark surveillance datasets were used for model training and evaluation:

- **UCSD Pedestrian Dataset (Peds1 and Peds2)** records low-density pedestrian pathways with random anomalies such as cyclists and wheelchairs.
- **CUHK Avenue Dataset** includes anomalies such as loitering, sudden running, and object throwing in a public pathway.
- **ShanghaiTech Dataset** contains diverse scenes from crowded passages to car roads with intricate anomalies.
- **Subway Dataset** provides real-world surveillance videos with scenarios such as fare evasion and walking in the opposite direction, recorded in station settings.

These datasets offered labeled normal and anomalous occurrences, allowing for strong model training using GAN-based models such as CVAD-GAN[17] and VALD-GAN[18]. Their scene diversity enables generalization to unobserved real-world situations

3.1.2 Real-World Dataset from Devaditya Technologies Pvt. Ltd. Indore

A key contribution to this research came from **Devaditya Technologies Pvt. Ltd.**, which donated **unprocessed 4K-resolution CCTV surveillance videos** recorded at Indian highway toll plazas. These videos constitute the primary dataset for training and testing the proposed system in a real-world deployment scenario.

The raw videos posed a number of computational difficulties as a result of their resolution and dimensions. To counteract this, a **custom automated preprocessing pipeline** was created. This pipeline effectively extracted frames and resized them based on the particular requirements of various deep learning models.

- For **CVAD-GAN**, where shape and motion are given more importance than fine color accuracy, the frames were resized to **160 × 160 pixels** and **converted to grayscale**.
- For **YOLOv8**, the same extracted frames were resized to **640 × 640 pixels**.

Figure 3.1: Original 4K resolution frame from surveillance video

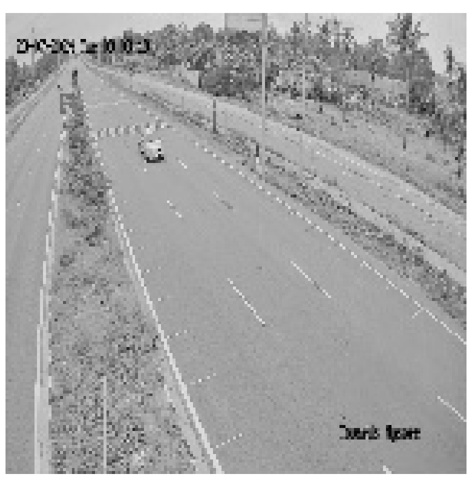


Figure 3.2: Resized grayscale frame (160×160) used for CVAD-GAN training

This multi-resolution strategy enabled parallel training of both anomaly detection and object detection models from a shared dataset with no degradation in performance and also no memory bottlenecks.

3.2 Data Preprocessing Pipeline

The following operations of preprocessing were routinely performed:

- **Frame Extraction**: Fixed-rate keyframes were extracted to have uniform temporal sampling.
- **Resolution Standardization**: Frames were resized to 160×160 for CVAD-GAN and to 640×640 for YOLOv8.
- **Format Conversion**: YOLOv8 bounding boxes were reshaped to normalized format: (center_x, center_y, width, height).
- **Augmentation**: Augmenting with random horizontal flipping, brightness modulation, and addition of Gaussian noise helped improve generalization.
- **Motion Feature Generation**: Optical flow and frame differencing were calculated to strengthen motion cues for GANs [**?**].

All these steps significantly improved the robustness and versatility of the proposed models.

3.3 Data Annotation

To enable **supervised training** for vehicle detection and axle classification tasks, approximately **2,500 frames** were **manually annotated** using a standardized labeling tool. Each image was tagged with bounding boxes around five classes:

- Car
- Bus
- Truck

- Motorcycle
- Wheel

Figure 3.3: Dataset Annotation

Manual annotation was particularly crucial for handling top-down camera angles, motion blur, and poor visibility during adverse weather conditions.

3.4 Summary

This chapter focuses on the acquisition and preprocessing of video datasets essential for training the object detection and anomaly detection models. It describes how high-resolution CCTV footage from Devaditya Technologies Pvt. Ltd. was collected and processed, including frame extraction, resizing, and grayscale conversion for CVAD-GAN training (160×160 resolution), and standardization to 640×640 resolution for YOLOv8. The chapter also covers manual annotation of over 2,500 frames with bounding boxes for vehicle types and axles, enabling supervised learning. Data augmentation techniques and motion feature extraction (like optical flow) were applied to improve model generalization and robustness across real-world highway surveillance scenarios.

Chapter 4

Video Incident Detection System using GAN

This chapter offers a deep generative method built on Constrained Video Anomaly Detection GAN (CVAD-GAN) to improve the anomaly detection features of the Vehicle Intelligence and Detection System (VIDS). The model is meant to learn the normal behavior of cars in highway scenes and flag deviations as anomalies. Unlike conventional classification techniques, this GAN-based method allows the system to detect subtle, unanticipated behaviors without depending on labeled anomaly data.

4.1 CVAD-GAN Architecture and Description

The CVAD-GAN[17] architecture consists of two main parts: a Generator and a Discriminator. It's particularly constructed for detecting spatiotemporal anomalies in video surveillance for learning to reconstruct typical traffic patterns and recognizing anomalies as possible deviations.

Each video frame is simplified by turning it into grayscale and shrinking it to a 160x160 pixel size. This makes the processing easier while keeping the important details intact. During training, the model only sees normal, everyday behavior. So, when it comes across something unusual during testing—like a car going the wrong way, a traffic jam, or an illegal stop—it struggles to recreate the scene accurately. This creates a noticeable error, which we use to flag the anomaly.

Figure 4.1 illustrates the overall CVAD-GAN framework used in this work.

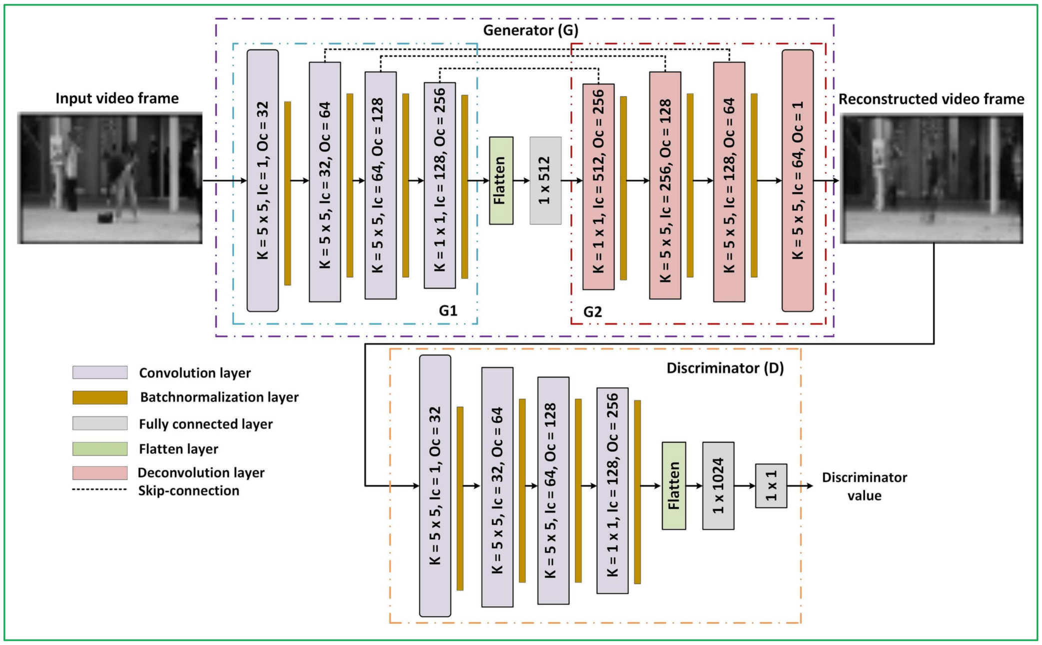


Figure 4.1: CVAD-GAN architecture for anomaly detection in highway surveillance[17]

The Generative Adversarial Network (GAN) is trained in an adversarial manner, where the generator learns to produce realistic-looking traffic frames from input sequences, aiming to deceive the discriminator. As training progresses, the generator gradually improves its ability to model and replicate typical motion patterns and structural characteristics observed in normal traffic flow.

4.2 Experimental Results

As discussed in Chapter 3, harnessing the visual data stream from highway sentinels – the surveillance cameras provided by Devaditya Technocrates Pvt. Ltd. – this study employed the CVAD-GAN model [17], guiding it through a meticulous learning process.The bedrock of this process was the scrupulous curation of training data; each and every video frame underwent careful human scrutiny to confirm its depiction of typical traffic flow. As a preliminary step, these visual snapshots were rendered in grayscale and resized to a uniform 160×160 pixel dimension. Subsequently, the model embarked on 150 iterative learning cycles, processing the data in manageable groups of 32 frames.

Upon completion of this learning phase, the model's capacity to discern the anomalous was put to the test, utilizing video sequences that presented a blend of both routine and unusual traffic occurrences. The system showcased its adeptness by effectively identifying several forms of non-standard activity, notably:

- Situations where vehicles braked unexpectedly to a standstill amidst the flow of moving traffic.
- Instances of vehicles navigating onto the highway from the incorrect direction.
- Scenarios where vehicles remained stationary for durations exceeding typical waiting times.

To provide a quantitative measure of the model's effectiveness, we utilized several key performance indicators:

- **Reconstruction Error**: CVAD-GAN effectively distinguishes between normal and anomalous video frames by leveraging constrained generation. The generator reconstructs normal frames with high fidelity while failing to do so for anomalous content, which results in visually and semantically distinguishable differences.
- **Area Under the ROC Curve (AUC)**: When evaluated on benchmark datasets like UCSD Ped2 and CUHK Avenue, CVAD-GAN achieves competitive Area Under the ROC Curve (AUC) scores. Specifically, it reports an AUC of 97.8% on UCSD Ped2 and 94.0% on CUHK Avenue, outperforming or matching several state-of-the-art methods in anomaly detection.

Figure 4.2 offers a visual example of a scenario where the model struggled to accurately recreate an anomalous frame, leading to a substantial reconstruction error and consequently, a successful detection of the anomaly.

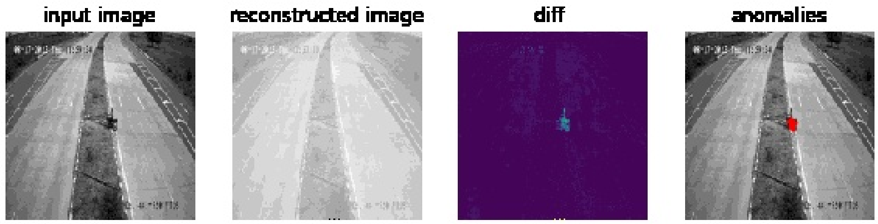


Figure 4.2: Example of CVAD-GAN output: input (left), reconstruction (middle), and error map (right) for an anomalous frame

These findings collectively highlight the potential of the CVAD-GAN framework for the unsupervised detection of anomalies within highway surveillance contexts. The model's capacity to pinpoint critical, unusual occurrences – often overlooked by more traditional, rule-based approaches – underscores its value in bolstering traffic monitoring capabilities and enhancing overall road safety.

4.2.1 Evaluation Metrics

Following the evaluation protocols adopted in prior works [17], we assess anomaly detection performance at the frame level. The primary evaluation tool is the Receiver Operating Characteristic (ROC) curve, which measures the discriminative capacity of the model. The ROC curve plots the True Positive Rate (TPR) against the False Positive Rate (FPR), calculated as follows:

$$\text{TPR} = \frac{TP}{TP + FN} \tag{4.1}$$

$$\text{FPR} = \frac{FP}{FP + TN} \tag{4.2}$$

Here, TP, FN, FP, and TN denote the counts of true positives, false negatives, false positives, and true negatives, respectively. In addition to the ROC curve, we report the Equal Error Rate (EER) and the Area Under the ROC Curve (AUC) for comprehensive performance evaluation. Furthermore, we compute the Area Under the Precision-Recall Curve (AU-PR), especially beneficial for imbalanced datasets. AU-PR is given by:

$$\text{AU-PR} = \int_0^1 \text{Precision}(R)\, dR \tag{4.3}$$

where R denotes the recall, and precision is the proportion of true positives among all predicted positives. A higher AU-PR implies stronger anomaly detection capabilities.

4.2.2 Implementation Details

All experiments were performed in Python using TensorFlow 2.13 on an NVIDIA Tesla V100 GPU. The hyperparameters used in the batch normalization layers include a small constant $\epsilon = 10^{-6}$ to ensure numerical stability and a decay factor of 0.9 for updating the moving averages. The weighting coefficients α_1 and α_2, which balance different loss components during training, are selected within the range $[0, 1]$. Empirical evaluations demonstrate that the model yields optimal performance when $\alpha_1 = 0.6$ and $\alpha_2 = 0.4$. These values of α_1 and α_2 are explicitly selected based on the accuracy of the model, as shown in Table 4.1.

Table 4.1: AUC values for different α_1 values on Peds2, Avenue, and ShanghaiTech datasets. For the Subway-Entry and Subway-Exit datasets, the number of detected anomaly events is reported.

α_1	Peds2 (%)	Avenue (%)	ShanghaiTech (%)	Subway-Entry	Subway-Exit
0.0	94.1	92.6	73.9	51	14
0.2	95.3	93.0	75.1	56	15
0.4	97.5	93.5	75.8	60	17
0.6	97.8	94.0	76.2	63	19
0.8	96.6	93.7	75.8	59	18
1.0	95.3	93.1	74.9	54	16

The discriminator outputs a confidence score in the range [0, 1] for each frame. Threshold values for anomaly detection were set based on dataset-specific tuning: 0.84 (Peds1), 0.53 (Peds2), 0.77 (Avenue), 0.73 (ShanghaiTech), 0.69 (Subway-Entry), and 0.51 (Subway-Exit). Epochs for training also varied: 100 (Peds1), 97 (Peds2), 84 (Avenue), 100 (ShanghaiTech), 60 (Subway-Entry), and 70 (Subway-Exit).

Contextual Performance

To provide a broader understanding of its capabilities, the original CVAD-GAN research reports the following performance metrics when applied to established benchmark datasets:

- **UCSD Peds1**: AUC of 98.0%
- **UCSD Peds2**: AUC of 97.8%
- **CUHK Avenue**: AUC of 94.0%
- **ShanghaiTech**: AUC of 76.2%

Source: [17]

While our specific highway surveillance data presents its own distinct set of challenges, the model's achieved performance remains notably competitive, illustrating its adaptability and potential across varied operational environments.

4.3 Summary

This chapter presents an anomaly detection system based on deep learning with the Constrained Video Anomaly Detection GAN (CVAD-GAN) to improve the Vehicle Intelligence and Detection System (VIDS). The structure includes a generator–discriminator pair trained solely on normal traffic patterns to detect anomalies based on reconstruction errors. The generator, as a convolutional autoencoder, reconstructs input frames (grayscale and resized to 160×160) while the discriminator separates real and generated frames. When it comes across abnormal scenes like inappropriate driving or long stops, the model does not reconstruct properly, thus indicating anomalies. Its effectiveness was proved with actual CCTV footage from Devaditya Technologies and benchmark datasets such as UCSD and CUHK Avenue with high AUC scores of 97.8%. Metrics like ROC, AUC, EER, and AU-PR were utilized to measure performance, and multiple hyperparameters (such as α_1 and α_2) were also optimized by tuning them. The results validated CVAD-GAN's ability to identify intricate anomalies under real-world highway monitoring scenarios.

Chapter 5

Deep Learning-Based Vehicle Analysis and Monitoring

This chapter describes the modules required to develop an intelligent highway traffic surveillance system based on real-time video analytics and deep learning. Section 5.1 introduces the Advance Traffic Counter and Classifier module that employs YOLOv8 for multi-class vehicle recognition, axle-based classification, and speed estimation. Section 5.1.1 explains the collection of real data using 4K CCTV video from Devaditya Technologies Pvt. Ltd. Section 5.1.2 discusses the hand annotation of more than 2,500 frames with bounding boxes for wheels and vehicles. Section 5.1.3 discusses multi-axle classification with visible cues and data augmentation to cope with occlusion and partial visibility. Section 5.1.4 describes the mechanism of real-time speed detection using ROI lines and frame-based motion analysis. Lastly, Section 5.2 proposes anomaly detection with CVAD-GAN that detects spatio-temporal irregularities in traffic patterns, providing strong surveillance even in difficult visual conditions.

5.1 Advance Traffic Counter and Classifier

Advance Traffic Counter and Classifier (ATCC) is one of the core modules Advance Traffic Counter and Classifier (ATCC) is a key module developed to automate real-time highway surveillance tasks such as vehicle detection, classification, counting, and speed estimation. It utilizes the YOLOv8 object detection model to accurately identify various vehicle types—cars, trucks, buses, and motorcycles—and further classifies heavy vehicles based on axle count, which is essential for tolling, traffic analytics, and regulatory compliance.

The system architecture, illustrated in Figure 5.1, includes five major modules:

- **Video Acquisition Module:** Collects real-time video streams from surveillance cameras installed at highway toll booths or strategic traffic points.

- **Object Detection and Classification Module:** Uses the YOLOv8 model to detect vehicles in each frame and classify them based on their type (car, truck, bus, etc.).

- **Multi-Axle Classification Module:** Further classifies heavy vehicles into categories based on the number of axles, which is crucial for toll calculation and regulatory enforcement.

- **Speed Estimation Module:** Calculates vehicle speed using the time difference between two frames and the known distance between two reference points on the road. This module also performs vehicle counting and motion tracking to support comprehensive traffic monitoring.

- **Data Storage and Visualization Module:** Records processed data in structured formats and presents it through a dashboard for real-time monitoring.

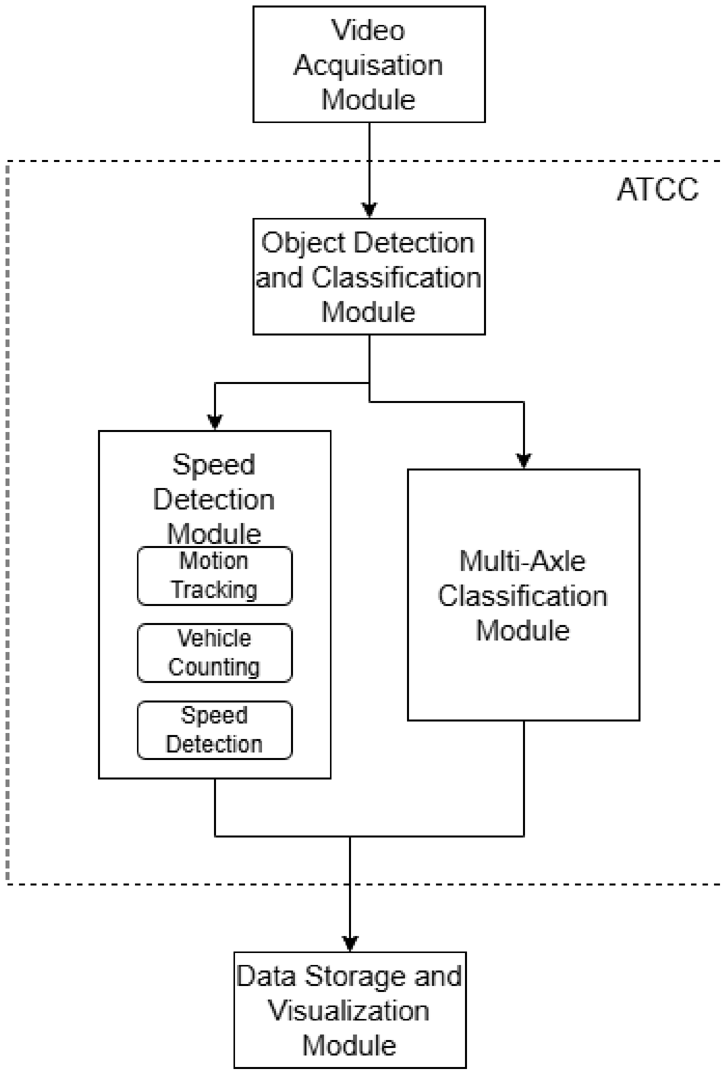


Figure 5.1: System architecture of the proposed Integrated Traffic Monitoring System

5.1.1 Real Data Collection

.

The surveillance data used here was obtained from Devaditya Technologies Pvt. Ltd. and comprised raw 4K video recordings taken at highway toll booths (see Section 3.1, Chapter 3). Due to the computational expense of handling high-resolution video, an automated pipeline was used for extracting efficient keyframes for downstream processing (see Section 3.2). These frames constituted the base input for training the anomaly detection module and the object classification system.

To ready data for training CVAD-GAN, every frame extracted was resized to 160 × 160 pixels and made grayscale to highlight motion and structural patterns rather than color (as described in Section 3.4). This simplified the computational overhead while maintaining fidelity for generative reconstruction.

Concurrently, the object detection and axle classification modules—YOLOv8-based—utilized the same dataset, downsized to 640 × 640 pixels to meet model input specifications (Section 3.3). The double-resolution preprocessing ensured that the input formats optimized both GAN-based and CNN-based models to allow for efficient and scalable training on real-world and benchmark datasets.

Figure 5.2: Original 4K frame from video

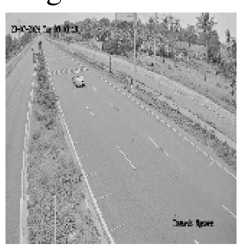


Figure 5.3: Resized grayscale frame used for CVAD-GAN training

5.1.2 Data Annotation

Approximately 2,500 frames were manually annotated to enable supervised learning for both vehicle detection and axle classification tasks. Using a standard annotation tool, each frame was labeled with bounding boxes around key vehicle types and components, such as wheels. The annotation process included five primary classes: `wheel`, `truck`, `motorcycle`,`car` and `bus`. These annotated frames serve as the ground truth necessary for training and evaluating the accuracy of both the axle classification and speed detection models.

An example of an annotated frame is shown in Figure 5.4, illustrating how different vehicle types and wheels are labeled within the scene.

Figure 5.4: Dataset Annotation

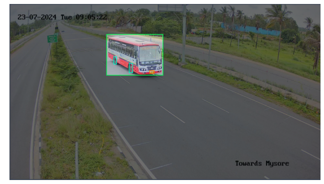
5.1.3 Extended Multi-Axle Vehicle Classification

Since toll rates are frequently correlated with axle count, it is essential for automated toll collection systems to accurately determine the number of axles on heavy vehicles. Multi-axle classification necessitates a more thorough examination of vehicle structure, particularly the axle configuration, in contrast to basic vehicle classification, which only differentiates between general types like trucks or buses.

This was addressed by manually annotating frames taken from surveillance videos with particular axle-based labels, like `2-Axle vehicle`, `3-Axle vehicle`, and `4-Axle vehicle` etc. The labeling procedure was guided by visual cues such as axle spacing, trailer extensions, and dual rear wheels. Figure 5.5 provides an example of annotated axle-based classifications.

(a) 2-Axle

(b) 3-Axle



(c) 4-Axle

(d) 4-Axle

Figure 5.5: Examples of multi-axle vehicle classification

YOLOv8s was selected for the classification task because of its effectiveness in striking a balance between detection speed and accuracy. For training and validation, the dataset was divided in an 80/20 ratio. Data augmentation techniques, such as random cropping, brightness adjustments, and horizontal flipping, were used to enhance generalization and simulate real-world driving conditions.

Occlusion was a major obstacle in this process; many vehicle frames only partially exposed their wheels, especially the far-side axles. Because of this, counting became challenging and possibly erroneous. In order to counteract this, the dataset was expanded to include more samples from different camera angles, which assisted the model in learning more resilient and viewpoint-invariant axle features.

The trained model's output contains structured detection results that are saved in JSON format, clearly encoding the type and axle classification of each vehicle. Real-time, automated, and equitable toll fee assignment based on axle count is made possible by the direct integration of these metadata files into tolling systems.

The output from the YOLOv8s model is shown in Figure 5.6, which shows how the system recognizes and categorizes cars in real-time based on axle configuration.

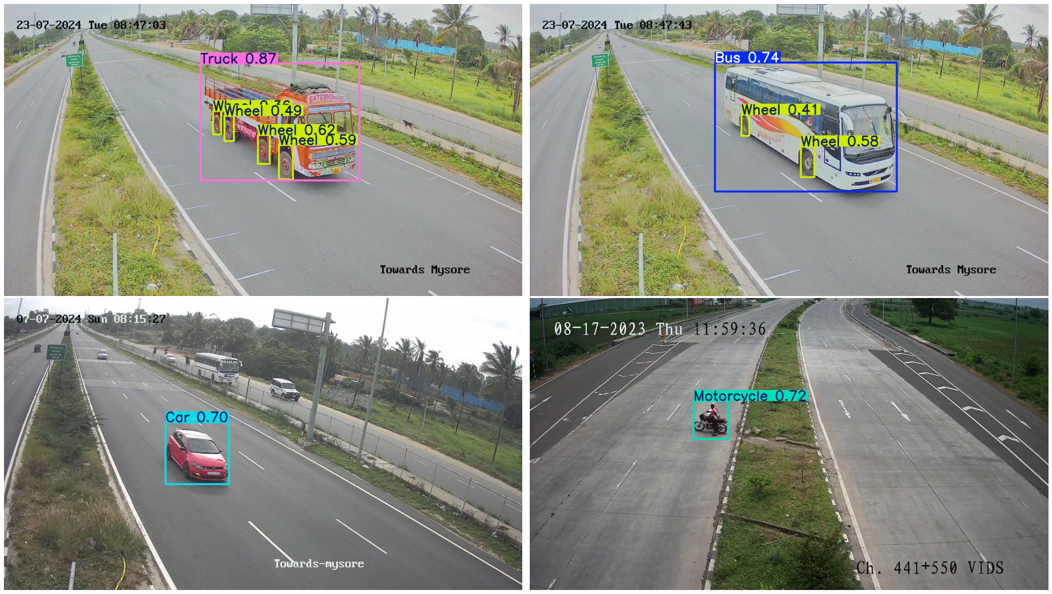


Figure 5.6: YOLOv8s detection output showing vehicle class and axle classification overlay

Challenges in Multi-Axle Vehicle Detection

Identifying and categorizing multi-axle vehicles from surveillance footage poses a number of practical difficulties that can have a big impact on automated systems' dependability and performance. These difficulties consist of:

- **Top-Down Camera View**: Because of the restricted visibility of the vehicle's side profile, counting wheels becomes challenging. The wheels, particularly those on the far side, are frequently obscured by a top-down view, rendering axle estimation inaccurate.

- **Poor Visibility as a Result of Weather**: Rain, fog, and dim lighting make photos less clear. In these circumstances, wheels are frequently not easily visible, making it more difficult to classify cars according to axle count.

- **Motion Blur in Rapidly Moving Vehicles:** Vehicles traveling at high speeds frequently produce motion-blurred frames, which reduce the detail in wheel regions and can result in missed axles or inaccurate detection.

- **Low Contrast between Wheels and Vehicle Body:** The detection model finds it challenging to discern the wheel boundaries when the wheel and the vehicle body have similar colors or textures, which leads to incorrect axle classification.

5.1.4 Real-Time Speed Detection

Because it ensures that vehicles follow traffic laws and promotes road safety, speed monitoring is essential to highway surveillance systems. In this study, a real-time, lightweight speed detection module was created to calculate vehicle velocities straight from CCTV footage. It is noteworthy that the system functions independently of external devices like embedded sensors, radar, or GPS.

The basic distance-over-time formula serves as the foundation for the system's operation. Two virtual lines, a *entry line* and a *exit line*, are used to define a Region of Interest (ROI) within each video feed. The current frame number is recorded when a car crosses the entry line. Another frame number is taken after the car crosses the exit line. The time required is calculated by multiplying the number of frames between these two events by the frame interval, which is obtained from the frame rate of the video. This makes it possible to calculate the vehicle's speed precisely, especially when combined with the known physical distance between the two lines.

Road markings or pre-established reference points that are visible in the video are used for camera calibration in order to guarantee accuracy. These calibrated distances aid in converting motion at the pixel level into actual units like meters per second (m/s) or kilometers per hour (km/h).

The speed detection module's final output is shown in Figure 5.7. It displays a detected vehicle on the video frame along with its defined ROI lines, speed value overlay, and bounding box.

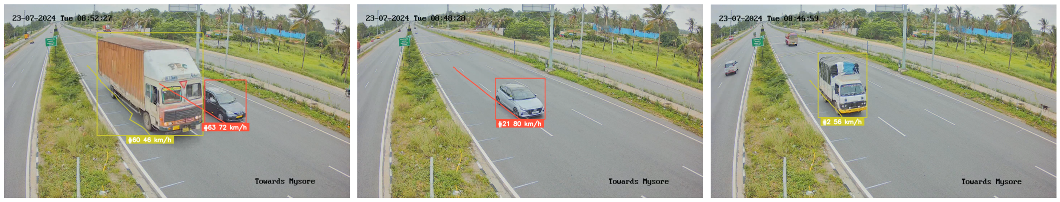


Figure 5.7: Real-time speed detection output showing vehicle bounding box, and estimated speed

This method provides a scalable and affordable solution that can be implemented in the real world at highway checkpoints and toll plazas. The system can be used to generate traffic statistics, inform dynamic toll pricing systems, and enforce speed limits.

Challenges in Speed Detection

While the speed detection module functions effectively in normal circumstances, real-world testing revealed a number of difficulties.

Poor visibility in inclement weather, such as fog, rain, or at night, was one of the main challenges. In these situations, cars frequently appeared faded or partially obscured, making it challenging to follow them precisely between the entry and exit lines.

Motion blur was another common problem, particularly with fast-moving cars. The blur reduced the accuracy of frame-based speed calculations by making bounding boxes unstable. Vehicles crossing the ROI in a small number of frames, which left little time for precise tracking, made this issue more noticeable.

Another problem was vehicle occlusion. The system frequently failed to identify or detect smaller vehicles until they were fully visible again when they were momentarily obscured by larger ones (like buses or trucks).

Last but not least, the intense glare from car headlights and reflections from damp road surfaces made nighttime surveillance difficult. Occasionally, these visual artifacts resulted in unsuccessful detections or false positives.

A multi-camera setup, thermal imaging, or adaptive brightness filtering may be necessary in the future to fully address these limitations, even though data augmentation and post-processing techniques somewhat increased the system's robustness.

5.2 Integration of Anomaly Detection Module with ATCC

The integration of the Constrained Video Anomaly Detection Generative Adversarial Network (CVAD-GAN) with the Advance Traffic Counter and Classifier (ATCC) gives strength to the system's operation through the integration of efficient anomaly detection in highway traffic monitoring videos. Although the ATCC module detects vehicles, classifies them, counts the axles, and estimates the speed in real time with high efficacy, CVAD-GAN refines it through the investigation of spatio-temporal patterns for detecting anomalous or irregular events within the traffic flow.

CVAD-GAN processes the preprocessed gray-scale frames extracted from the same CCTV clips utilized by ATCC, reduced to $160X160$ pixels to preserve vital structural features with optimized computation speed. Sharing a common data pipeline guarantees smooth integration between the two modules. The CVAD-GAN generator learns to encode normal traffic patterns by reconstructing input frames, while the discriminator learns to classify between real and generated frames in order to allow the system to identify anomalies like accidents, vehicles that have broken down, or unusual traffic congestion.

Through the integration of CVAD-GAN's anomaly detection output with ATCC-generated structured vehicle metadata like vehicle type, number of units, and velocity, the system as a whole is an end-to-end traffic monitoring system. Such an integrated system gives intelligent transportation systems not just the ability to monitor regular traffic parameters but also the capability to detect and respond to serious events beforehand, thereby improving traffic safety and management.

The architecture supports real-time processing through parallel execution of the ATCC and CVAD-GAN modules, with aggregated results visualized through a single dashboard. This integration offers traffic authorities both quantitative vehicle information and qualitative anomaly alerts to facilitate timely decision-making and effective traffic control measures.

Figure 5.8: Anomaly visualization in the Flask application dashboard indicating detected unusual traffic events.

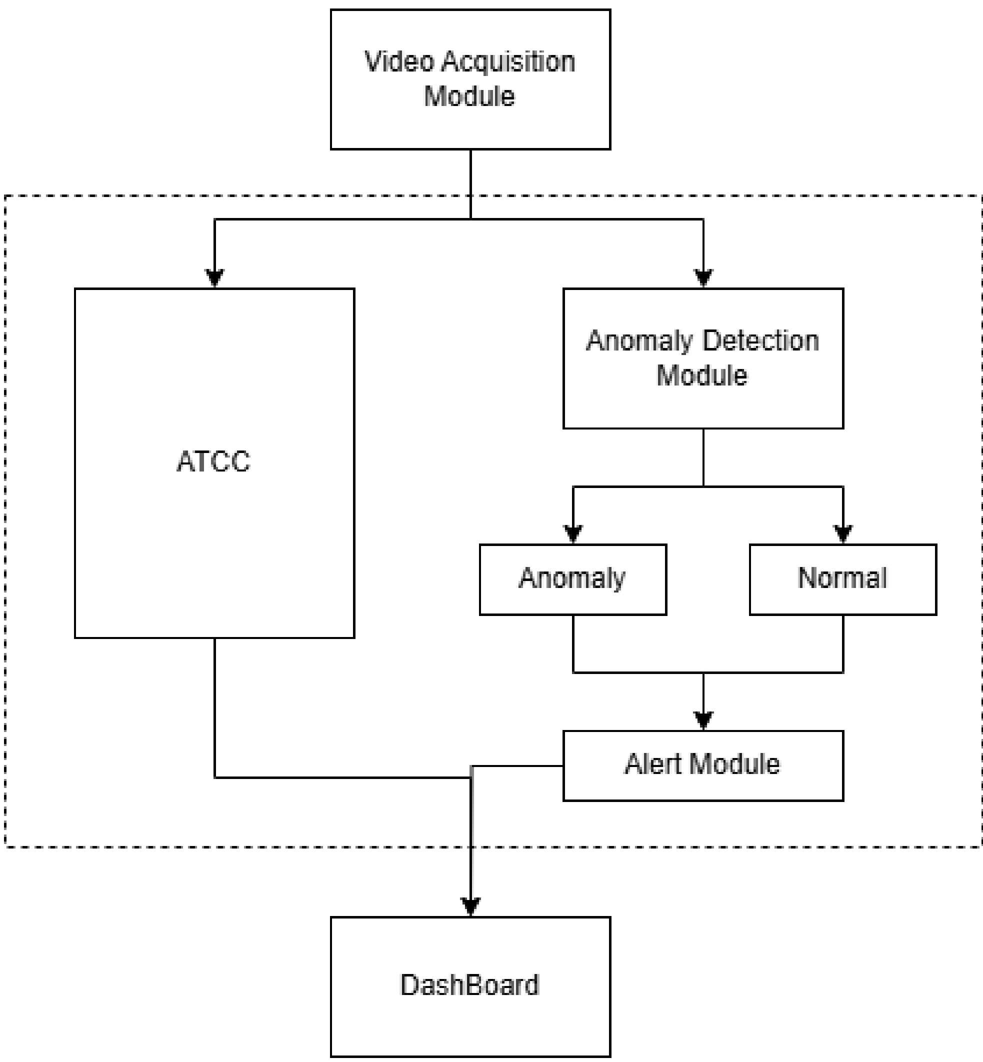


Figure 5.9: System architecture integrating ATCC with anomaly detection module

The system architecture shown in Figure 5.9 displays an all-encompassing structure that combines the *Automatic Traffic Counter and Classifier (ATCC)* with an *Anomaly Detection Module* to advance highway surveillance and incident detection. The pipeline starts with the *Video Acquisition Module*, which captures real-time images from roadside surveillance cameras. The visual data is then passed to the *ATCC*, which processes the video to extract traffic statistics such as vehicle count, class, and speed. At the same time, the same input goes to the *Anomaly Detection Module*, which examines spatio-temporal patterns in order to identify anomalies like abrupt stops, wrong-way travel, or stopped vehicles. Depending on the analysis, the system labels the data as either "Normal" or "Anomaly". In the event of anomalies, alarms are raised through the *Alert Module*, providing timely alerting of important events. All the outputs, such as traffic statistics and anomaly warnings, are then displayed through a centralized *Dashboard* to aid in effective traffic management and decision-making. This architecture has the benefits of modularity, scalability, and real-time responsiveness, which are fundamental requirements for intelligent transportation systems.

Conclusion

This chapter presents an integrated deep learning-based system for highway vehicle analysis and monitoring, comprising modules for data acquisition, annotation, vehicle classification, axle counting, speed estimation, and anomaly detection. The real-time traffic counter and classifier system begins with collecting high-resolution 4K CCTV footage from Devaditya Technologies Pvt. Ltd., followed by annotation of over 2,500 frames into categories such as car, truck, motorcycle, and wheel for training YOLOv8. This model enables accurate multi-axle vehicle classification under varied conditions through augmentation techniques and improved viewpoint generalization. A real-time speed detection mechanism is implemented using virtual ROI lines and timestamped frame differences, eliminating the need for external sensors. The anomaly detection sub-system utilizes CVAD-GAN, which learns normal traffic behavior through generative reconstruction and flags anomalies—such as accidents or wrong-way driving—based on deviations between the original and generated frames.

Chapter 6

Future Work And Limitations

The proposed highway traffic monitoring system shows robust performance in vehicle detection, axle classification, speed estimation, and anomaly detection, there are certain limitations. One major limitation lies in the lack of a real-time alert mechanism that can directly alert authorities at once when anomalies like wrong-way driving or illegal stops occur. Moreover, while the system effectively detects anomalies through CVAD-GAN, it does not yet classify the anomaly type, for instance, accidents, obstructions, or sudden stops, which prevents contextual interpretation of the event being detected.

The system also does not have number plate recognition, which limits its capability to associate detected events with particular vehicles. This is particularly significant for enforcing toll offenses or detecting unusual behavior vehicles. Additionally, GAN-based models though being effective, can be trained using normal data and might fail to generalize to unseen or infrequent anomalies, especially under diverse environmental and traffic conditions. Environmental elements such as low lighting, glare, rain, and motion blur can affect detection performance, particularly in axle counting and speed measurement. Moreover, the system needs to be calibrated manually for region-of-interest (ROI) configuration in vehicle tracking and speed estimation, lowering flexibility at new deployment locations. The computational overhead of the GAN models can also be a potential concern for real-time execution on edge devices without optimization.

In the future, the system can be further enhanced by incorporating a real-time alert module that is capable of pushing SMS, email, or web notifications to highway officials on detecting anomalies. Adding automatic number plate recognition (ANPR) will further strengthen enforcement by linking certain anomalies with vehicle identities. To enhance model generalizability to varying contexts and illumination conditions, approaches like domain adaptation and low-rank adapter (LoRA)-based fine-tuning can be used on the GANs. Classifying the anomalies identified into types would also be a focus of future work, allowing response prioritization based on event severity. Calibration of ROI using lane marker detection automation would enhance scalability. Moreover, improving the speed estimation module with multi-camera arrangements or depth data can overcome present constraints with occlusion and motion blur. Lastly, quantization or model compression of GAN inference can make it more viable to deploy on edge computing hardware, and a central dashboard with browsable event logs may offer complete traffic insight for the authorities.

Bibliography

[1] Abati, D., Porrello, A., Calderara, S., and Cucchiara, R. Latent space autoregression for novelty detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (2019), pp. 481–490.

[2] Cong, Y., Yuan, J., and Liu, J. Sparse reconstruction cost for abnormal event detection. In *CVPR 2011* (2011), IEEE, pp. 3449–3456.

[3] Courville, A., and Bengio, Y. Generative adversarial nets. *Advanc in Neural 27* (2014), 3.

[4] Golan, I., and El-Yaniv, R. Deep anomaly detection using geometric transformations. *Advances in neural information processing systems 31* (2018).

[5] Gong, D., Liu, L., Le, V., Saha, B., Mansour, M. R., Venkatesh, S., and Hengel, A. v. d. Memorizing normality to detect anomaly: Memory-augmented deep autoencoder for unsupervised anomaly detection. In *Proceedings of the IEEE/CVF international conference on computer vision* (2019), pp. 1705–1714.

[6] Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. Generative adversarial nets. *Advances in neural information processing systems 27* (2014).

[7] Han, X., Chen, X., and Liu, L.-P. Gan ensemble for anomaly detection. In *Proceedings of the AAAI Conference on Artificial Intelligence* (2021), vol. 35, pp. 4090–4097.

[8] Hu, E. J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., Chen, W., et al. Lora: Low-rank adaptation of large language models. *ICLR 1*, 2 (2022), 3.

[9] Karpathy, A., Toderici, G., Shetty, S., Leung, T., Sukthankar, R., and Fei-Fei, L. Large-scale video classification with convolutional neural networks. In *2014 IEEE Conference on Computer Vision and Pattern Recognition* (2014), pp. 1725–1732.

[10] Lee, S., Kim, H. G., and Ro, Y. M. Stan: Spatio-temporal adversarial networks for abnormal event detection. In *2018 IEEE international conference on acoustics, speech and signal processing (ICASSP)* (2018), IEEE, pp. 1323–1327.

[11] Li, N., and Chang, F. Video anomaly detection and localization via multivariate gaussian fully convolution adversarial autoencoder. *Neurocomputing 369* (2019), 92–105.

[12] Li, Q., Yang, R., Xiao, F., Bhanu, B., and Zhang, F. Attention-based anomaly detection in multi-view surveillance videos. *Knowledge-Based Systems 252* (2022), 109348.

[13] Li, W., Mahadevan, V., and Vasconcelos, N. Anomaly detection and localization in crowded scenes. *IEEE Transactions on Pattern Analysis and Machine Intelligence 36*, 1 (2014), 18–32.

[14] Lu, C., Shi, J., and Jia, J. Abnormal event detection at 150 fps in matlab. In *Proceedings of the IEEE international conference on computer vision* (2013), pp. 2720–2727.

[15] Mao, X., Li, Q., Xie, H., Lau, R. Y., Wang, Z., and Paul Smolley, S. Least squares generative adversarial networks. In *Proceedings of the IEEE international conference on computer vision* (2017), pp. 2794–2802.

[16] Mehran, R., Oyama, A., and Shah, M. Abnormal crowd behavior detection using social force model. In *2009 IEEE conference on computer vision and pattern*

recognition (2009), IEEE, pp. 935–942.

[17] SINGH, R., SETHI, A., SAINI, K., SAURAV, S., TIWARI, A., AND SINGH, S. Cvadgan: Constrained video anomaly detection via generative adversarial network. *Image and Vision Computing 143* (2024), 104950.

[18] SINGH, R., SETHI, A., SAINI, K., SAURAV, S., TIWARI, A., AND SINGH, S. Valdgan: video anomaly detection using latent discriminator augmented gan. *Signal, Image and Video Processing 18*, 1 (2024), 821–831.

[19] WOJKE, N., BEWLEY, A., AND PAULUS, D. Simple online and realtime tracking with a deep association metric. In *2017 IEEE International Conference on Image Processing (ICIP)* (2017), pp. 3645–3649.

[20] WU, P., LIU, J., AND SHEN, F. A deep one-class neural network for anomalous event detection in complex scenes. *IEEE Transactions on Neural Networks and Learning Systems 31*, 7 (2020), 2609–2622.

[21] YU, J., LEE, Y., YOW, K. C., JEON, M., AND PEDRYCZ, W. Abnormal event detection and localization via adversarial event prediction. *IEEE transactions on neural networks and learning systems 33*, 8 (2021), 3572–3586.

[22] YU, J.-H., MOON, J.-H., AND SOHN, K.-A. Attention-guided residual frame learning for video anomaly detection. *Multimedia Tools and Applications 82*, 8 (2023), 12099–12116.

Appendix A Content

Appendix B Content