




Article

Hybrid Machine Learning for Automated Road Safety Inspection of Auckland Harbour Bridge

Munish Rathee ^{1,*} , Boris Bačić ^{1,*}  and Maryam Doborjeh ^{1,2} 

¹ School of Engineering, Computer and Mathematical Sciences, Auckland University of Technology, Auckland 1010, New Zealand; mgholami@aut.ac.nz or maryam.gholami.doborjeh@aut.ac.nz

² Knowledge Engineering and Discovery Research Innovation, Auckland University of Technology, Auckland 1010, New Zealand

* Correspondence: munishonthenet@gmail.com or munish.rathee@aut.ac.nz (M.R.); boris.bacic@aut.ac.nz (B.B.)

Abstract: The Auckland Harbour Bridge (AHB) utilises a movable concrete barrier (MCB) to regulate the uneven bidirectional flow of daily traffic. In addition to the risk of human error during regular visual inspections, staff members inspecting the MCB work in diverse weather and light conditions, exerting themselves in ergonomically unhealthy inspection postures with the added weight of protection gear to mitigate risks, e.g., flying debris. To augment visual inspections of an MCB using computer vision technology, this study introduces a hybrid deep learning solution that combines kernel manipulation with custom transfer learning strategies. The video data recordings were captured in diverse light and weather conditions (under the safety supervision of industry experts) involving a high-speed (120 fps) camera system attached to an MCB transfer vehicle. Before identifying a safety hazard, e.g., the unsafe position of a pin connecting two 750 kg concrete segments of the MCB, a multi-stage preprocessing of the spatiotemporal region of interest (ROI) involves a rolling window before identifying the video frames containing diagnostic information. This study utilises the ResNet-50 architecture, enhanced with 3D convolutions, within the STENet framework to capture and analyse spatiotemporal data, facilitating real-time surveillance of the Auckland Harbour Bridge (AHB). Considering the sparse nature of safety anomalies, the initial peer-reviewed binary classification results (82.6%) for safe and unsafe (intervention-required) scenarios were improved to 93.6% by incorporating synthetic data, expert feedback, and retraining the model. This adaptation allowed for the optimised detection of false positives and false negatives. In the future, we aim to extend anomaly detection methods to various infrastructure inspections, enhancing urban resilience, transport efficiency and safety.

Keywords: anomaly detection; structural damage detection; traffic safety; computer vision; machine learning; deep learning; transfer learning; ARDAD



Citation: Rathee, M.; Bačić, B.; Doborjeh, M. Hybrid Machine Learning for Automated Road Safety Inspection of Auckland Harbour Bridge. *Electronics* **2024**, *13*, 3030. <https://doi.org/10.3390/electronics13153030>

Academic Editors: Jiwei Zhang and Shaozhang Niu

Received: 15 June 2024

Revised: 24 July 2024

Accepted: 29 July 2024

Published: 1 August 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The Auckland Harbour Bridge, spanning 1.2 km across Waitemata Harbour, was opened on 30 May 1959. Initially handling 11,205 vehicles daily, the bridge currently accommodates around 154,000 vehicles daily, with peaks over 200,000 due to public transport shifts [1]. The bridge supports rapid regional development, with a quadrupled North Shore population in the past 50 years. The New Zealand Transport Authority (NZTA), also known as Waka Kotahi, annually invests up to NZD 4 million in its maintenance and employs about 160 people for ongoing upgrades and maintenance. Like all crucial transportation infrastructure in New Zealand, the Auckland Harbour Bridge (AHB) faces significant maintenance challenges due to environmental and external factors. Such challenges contribute to the country's overall high maintenance costs for road infrastructure, which amount to as much as 1.1% of New Zealand's GDP [2].

This paper is organised as follows:

Section 1 presents an overview of the problem and the research question; Section 2 includes a comprehensive literature review on Automated Road Defect and Anomaly Detection; and Section 3 proposes a new methodology based on a hybrid deep learning solution to augment visual inspections using computer vision technology.

1.1. Background

Installed in 1990, the movable concrete barrier (MCB) on the AHB has enhanced rush-hour traffic flow and safety. Every weekday, the AHB utilises two Barrier Transfer Machines (BTMs) to adjust lane configurations, improving traffic flow during peak hours by moving 750 kg concrete blocks. The BTMs, essential for managing the daily tidal, were introduced in 2009 for 1.4 million NZD each. Despite the high cost, the BTMs do not have an automated surveillance system to ensure the integrity of the movable concrete barriers (MCBs). In the absence of an MCB barrier safety inspection system, NZTA staff must walk over a mile in hazardous conditions and amidst dangerous traffic to manually inspect and ensure the integrity of the MCB [3].

The reliability of the MCB system depends on the integrity of the metal pins connecting the barrier segments (Figure 1). The critical function of the pins is to secure barrier segments that regulate lane division based on traffic flow [3]. Malfunctioning or dislodged pins pose risks to traffic safety and impede the system's efficiency, leading to potential traffic disruptions and increased accident risks.

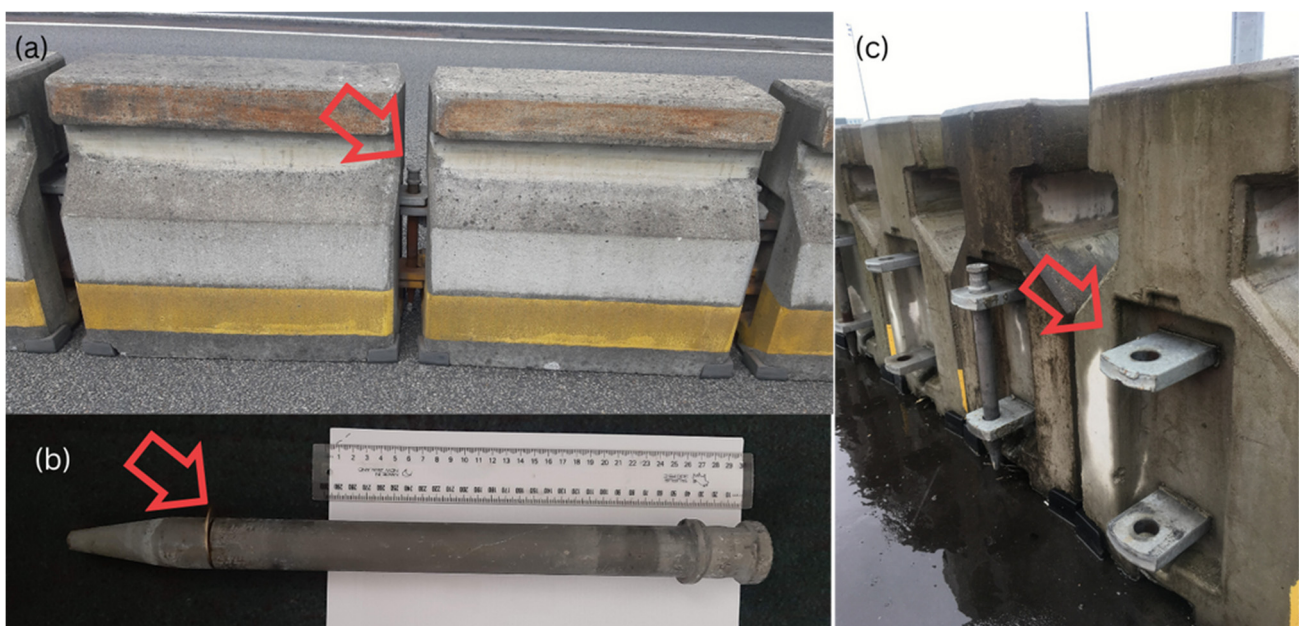


Figure 1. A movable concrete barrier system and its safety challenges. (a) Pin position requiring fixing, (b) metal pin with detachable safety ring, and (c) movable concrete barrier joints without metal pins.

Prior works in automated infrastructure inspection often fall short in dynamic and complex environments like the AHB [4–6]. This research aims to advance the field by incorporating a novel hybrid deep learning and spatiotemporal data analysis, allowing for a more accurate and reliable detection of safety anomalies in complex environments such as the AHB. This study employs spatiotemporal analysis, deployable AI algorithms, and semi-automated synthetic data generation to enhance traffic barrier monitoring, transforming research into practical, real-time anomaly detection solutions. The objective is to reduce the risks linked with manual inspections, enhance traffic safety (Figure 2), and boost the efficiency of barrier systems. Building on the Proof-of-Concept developed in 2019–2020 [7], this research utilises computer vision and deep learning techniques to au-

tomate the detection of metal pins in movable concrete barriers (MCBs) on the Auckland Harbour Bridge (AHB).



Figure 2. The scale of the challenges for manual inspection and fixing of the metal pin: (a) various locations of movable concrete barriers on and around the Auckland Harbour Bridge; (b) finding an unsafe pin and fixing it manually.

Hybrid machine-learning models, which integrate CNNs with other machine-learning techniques, can revolutionise inspections by enhancing the accuracy and reliability of detecting and classifying anomalous conditions [8,9]. The deep learning models trained on the synthetically enhanced dataset use a feature-based detection approach to analyse video frames for signs of pin displacement. Real-time monitoring capability is crucial for effective traffic management, especially during peak hours. Immediate alerts to bridge operators facilitate timely interventions, preventing potential safety issues from escalating into more severe incidents. Developing such an automated surveillance system advances civil engineering and traffic safety, offering a scalable, efficient solution to a longstanding safety challenge and setting a precedent for similar applications worldwide, potentially leading to the broader adoption of intelligent traffic management solutions in global urban settings [10].

This study developed a hybrid machine learning system for real-time, privacy-preserving anomaly detection in road safety inspections. This research's contributions are listed as follows:

- Produces a traffic safety analysis artefact scalable to scenarios in over 20+ countries and hundreds of similar traffic scenes [10] that employ movable concrete barriers.
- Introduces a semi-automated synthetic data generation method using a novel background cloning technique. The novel approach addresses data sparsity and enhances model training, with repeatability value for other computer vision case studies facing dataset balancing issues.
- Refines classification methods to balance false positives and negatives, improving detection accuracy from 82.6% (reported in earlier peer-reviewed research [11]) to 93.6%. Achieving this 11% increase in accuracy within complex traffic scenes characterised by chaotic backgrounds and lighting conditions underscores the artefact's viability for real-world applications.
- Successfully navigates hazardous traffic scenes for data collection by adhering to industry safety protocols. This repeatable approach provides a comprehensive blueprint for managing similar scenarios, ensuring stakeholder satisfaction and achieving sufficient data.

1.2. Research Questions and Modelling Concepts

This paper presents the development of automated solutions leveraging computer vision and deep learning to enhance traffic safety and operational efficiency by automating the inspection of metal pins in movable concrete barriers. This research explores visual sensors' potential to detect transport activity anomalies while maintaining privacy. Focused on the Auckland Harbour Bridge's movable barrier, our research automates safety screening and enhances anomaly detection for critical yet underrepresented classes like unsafe pin positions. Furthermore, classification techniques were refined to balance false positives and negatives, thereby improving the reliability and effectiveness of the traffic safety system.

1. Can relevant information from visual sensors be extracted for anomaly detection in transport activity while preserving privacy?
 - (a) What are the optimal methods for extracting vital information from visual media to detect transport activity anomalies without compromising privacy?
 - (b) To what degree can the safety screening of Auckland Harbour Bridge's movable barrier and the surrounding areas be automated using AI, CV, and DL methods?
2. How can synthetic data generation be streamlined to portray minority classes, such as unsafe pin positions, in anomaly detection tasks within traffic safety?
3. How can classification performance be honed to optimise an equilibrium between false positives and negatives from the early Proof-of-Concept (PoC) [7] while considering present and anticipated data scenarios?

2. Literature Review

Automated Road Defect and Anomaly Detection (ARDAD) uses computer vision, combining traditional and deep learning methods with unsupervised learning [12,13]. Traditional unsupervised methods, relying on datasets of normal conditions, often produce high false alarm rates in complex environments [14]. This research employs the Spatio-Temporal Enhanced Network (STENet) to address the problem of complex traffic scenes, which leverages temporal and spatial data for better generalisation and robust anomaly detection.

2.1. Auckland Harbour Bridge

The New Zealand Transport Agency (NZTA) manages the Auckland Harbour Bridge (AHB), an eight-lane motorway supporting over 200,000 vehicles daily. Installed in the 1990s, the movable concrete lane barrier (MCB) is essential for preventing crashes and optimising traffic flow during peak hours. The MCB system, consisting of 750 kg concrete blocks connected by metal pins, is prone to displacement due to traffic and ambient vibrations, posing significant risks [1,3,15]. Barrier Transfer Machines (BTMs) facilitate the movement of the MCB, typically adjusting lanes four times daily to manage traffic flow effectively.

Manual inspection of the pins is labour-intensive and occurs under poor ergonomic conditions, making it susceptible to human error. Despite the system's efficiency in altering traffic lanes swiftly—approximately 10 min for a one-kilometre section—the lack of automated pin inspection mechanisms within the existing NZD 1.4 million BTMs today may be seen as a significant oversight and an opportunity to apply CV in maintenance and safety protocols [1,15,16]. While effective in managing contraflow and heavy occupancy lanes, the movable concrete barrier (MCB) system requires continual manual monitoring to ensure its structural integrity and operational reliability.

2.2. Review of Surveys on ARDAD

The application of artificial intelligence (AI) and deep learning in road infrastructure anomaly and object detection has achieved significant breakthroughs over the past decade [5,12,13]. AI advancements have impacted domains such as autonomous driving,

face recognition, and personalised healthcare [17,18]. Deep learning techniques, particularly Convolutional Neural Networks (CNNs), have revolutionised image processing by automating segmentation, recognition, and reconstruction tasks [6]. CNNs efficiently process and classify image data through pooling and convolution, extracting features and reducing dimensionality [19–22]. Unlike traditional methods that rely on hand-engineered features, deep learning models learn feature representations directly from data. Architectures like R-CNN and its variants excel in object detection using region proposal methods to pre-selected areas of interest [23].

Moreover, transfer learning enhances model generalizability by applying models trained on one task to related but different tasks [24]. Transfer learning leverages pre-trained models to improve performance on new datasets, often by fine-tuning the final layers of networks like ResNet-50, which demonstrated remarkable success in the Visual Recognition Challenge [25]. The approach underscores deep learning models' versatility and adaptive capacity in handling diverse and complex visual data environments.

2.3. Video and Image Optimisation Techniques in ARDAD

Selecting an appropriate colour space in image segmentation is crucial and often application-specific, with no consensus on the best choice. Common colour spaces include RGB, LAB, CMY, XYZ, HSV, YCbCr, YIQ, YUV, and DHT [26,27]. For example, RGB is effective for bilirubin concentration changes (Equation (1)), while CMY and HSV excel in other tasks [28]. Advanced methods like one-dimensional histograms for automatic colour space selection [29] and two-dimensional histograms for improved segmentation [30] have been developed to optimise the process. Specific applications, such as YCbCr for face detection [31] and YIQ for satellite imagery [32], demonstrate tailored approaches. For monitoring the Auckland Harbour Bridge (AHB), the Y component (Equation (1)) from RGB to YCbCr transformation can enhance structural anomaly detection by focusing on luminance and minimising colour variations due to lighting or weather conditions.

$$Y = 0.299R + 0.587G + 0.114B \quad (1)$$

Techniques like histogram thresholding and SVM-based clustering have also been applied to enhance segmentation by structuring pixel data into coherent colour groups [33,34].

2.4. Advanced Techniques in Video and Image Analysis for ARDAD

In computer vision, object detection in videos involves recognising the movement of objects across multiple frames, utilising techniques like background subtraction, frame differencing, and optical flow. The Gaussian Mixture Model (GMM) is a prevalent method for background separation, enhancing foreground–background distinction by modelling pixel distributions with Gaussian mixtures [35–37]. Template matching, another practical approach, uses MATLAB 2022a functions like *normxcorr2* and *regionprops* to track objects in successive frames by identifying the peak of normalised cross-correlation [38]. Template matching, as represented by Equation (2), calculates the intensity-weighted centroid of an object using the following expression:

$$x_c = \frac{\sum_{i=1}^N x_i \cdot w_i}{\sum_{i=1}^N w_i} \quad (2)$$

where x_c is the centroid location, x_i is the pixel location, and w_i is the pixel intensity. Multiplying x_i by w_i in the numerator ensures that each pixel's location is weighted by its intensity, giving more importance to pixels with higher intensities in the centroid calculation.

Foreground–background separation techniques split a video into static backgrounds and dynamic foregrounds, using models that adapt to changes in the scene to detect motion and recognise objects [39,40]. Recent advancements in the GMM have introduced methods that handle complex scenes like traffic, improving the detection of slow-moving

or oversized vehicles [41]. Each pixel in the video sequence is modelled as a mixture of K Gaussian distributions. The probability of observing a pixel value X at time t is given by Equation (3).

$$P(X_t) = \sum_{k=1}^K \omega_{k,t} \cdot N(X_t | \mu_{k,t}, \Sigma_{k,t}) \quad (3)$$

Here, $\omega_{k,t}$ is the weight of the k^{th} Gaussian component at time t , N is the Gaussian distribution, $\mu_{k,t}$ is the mean of the k^{th} Gaussian, and $\Sigma_{k,t}$ is the covariance matrix of the k -th Gaussian. On the other hand, frame difference methods detect motion by comparing pixel differences from one frame to the next, a simple yet effective technique for identifying moving objects [42]. Optical flow techniques assess motion by analysing the pattern of apparent motion of objects, surfaces, and edges in a visual scene caused by the relative motion between an observer and the scene [43]. For example, Equation (4) synthesises motion detection using frame differences $|F_t - F_{t-1}|$ to identify changes and optical flow, represented by $\nabla I_t \cdot v_t$, to analyse motion patterns and speeds, with α adjusting their relative contributions.

$$\Delta M_t = |F_t - F_{t-1}| + \alpha \cdot \nabla I_t \cdot v_t \quad (4)$$

Object tracking in computer vision follows object detection, using colour, texture, shape, size, and orientation to track objects like vehicles or pedestrians across video frames. Robust tracking is essential across camera placements, lighting conditions, and cluttered scenes [44]. Techniques like the Kalman filter and particle filter are standard for tracking linear and non-linear motions [45]. Blob analysis tracks objects in binarised images based on features like area and bounding-box dimensions [46]. The Bayer filter pattern calculates the distance between object centroids, aiding in tracking and measuring vehicle velocity. The Kalman filter refines trajectory predictions by combining predicted and actual locations [47,48]. For non-linear scenarios, the Extended Kalman filter uses the Jacobian matrix for transition matrices (Equation (5)).

$$\hat{x}_{k|k} = \hat{x}_{k|k-1} + K_k (z_k - H_k \hat{x}_{k|k-1}) \quad (5)$$

In Equation (5), $\hat{x}_{k|k}$ is the updated state estimate at time k ; $\hat{x}_{k|k-1}$ is the predicted state estimate at time k , based on the estimate from time $k-1$; K_k is the Kalman gain, which dictates the blending of prediction with observation; z_k is the actual measurement at time k ; and H_k is the measurement matrix that maps the state space into the measurement space.

The particle filter enhances tracking in complex scenarios by sampling and resampling object features across frames, proving more effective in non-linear and cluttered environments than other methods [49,50]. It utilises a correlation particle filter to manage scale variations and feature interdependencies.

The mean shift algorithm, which uses a statistical colour model for tracking, iteratively converges to high-density areas in the colour space 8-point connectivity to locate objects [51]. The Adaptive Local Movement Model (ALMM) addresses movement by focusing on regional patches rather than the entire object, proving effective against occlusions and rapid movements [52].

Data collection and manual labelling in ARDAD systems can be resource-intensive and time-consuming. While deep learning approaches offer significant advantages in image and video processing due to their ability to outperform traditional methods, they require substantial datasets and computational power [53]. Alternatively, expert-driven feature extraction with traditional machine learning (ML) approaches can be effective with smaller datasets and less computational demand, making them suitable for initial target system designs [54]. Pretrained deep learning models facilitate automatic feature extraction and enable transfer learning, which adapts models to new functions without extensive computational resources [55] and catastrophic forgetting. The fusion of transfer learning with traditional machine learning enhances spatiotemporal classification by leveraging pre-trained models for efficient feature extraction and applying traditional models for sequence analysis, presenting an opportunity for improving classification accuracy (Figure 3).

Combining deep learning with traditional machine learning has been applied in various domains, such as traffic flow prediction and crop health monitoring [56,57].

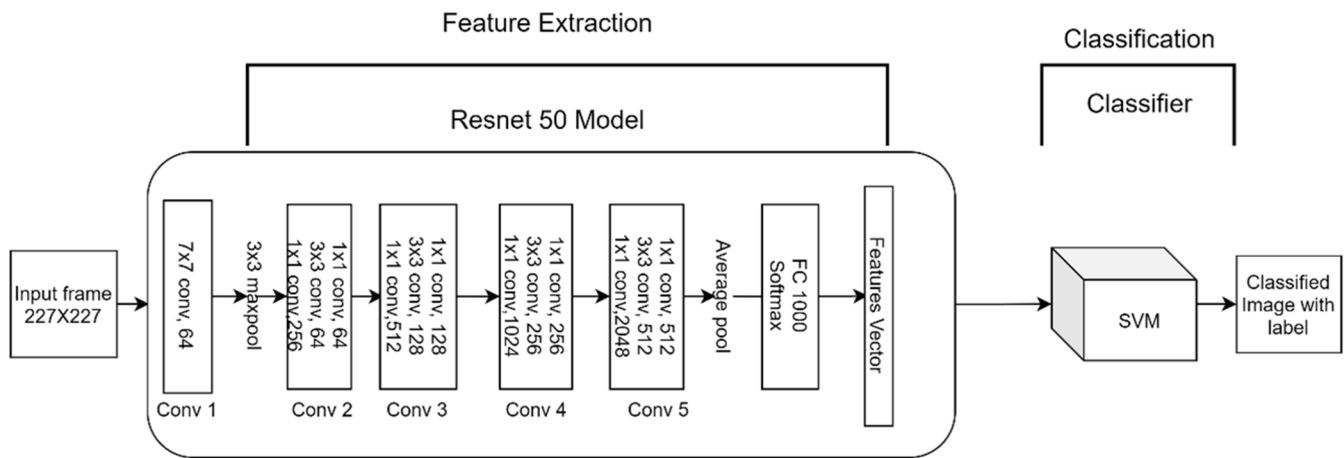


Figure 3. Integration of transfer learning with traditional ML: ResNet 50 extracts features from input images, which an SVM classifies. Integrating ResNet 50’s deep learning capabilities to generate feature space without relying on expert knowledge combined with a traditional classifier (e.g., SVM) presents the opportunity to enhance image classification performance.

The fusion of transfer learning with traditional machine learning (Figure 3) for spatiotemporal classification tasks is intended to replace the softmax() function (6), which is commonly used as the last layer of a neural network to convert high-dimensional feature vectors ($\in \mathbb{R}$) typically into a probability of possible outcomes, i.e., normalised for intended output class distribution.

$$\hat{y} = \text{softmax}(Wf_{\text{RNN}}(f_{\text{CNN}}(X)) + b) \quad (6)$$

where the loss function used in transfer learning is

$$L(\theta) = \frac{1}{N} \sum_{i=1}^N \ell(y_i, f(X_i; \theta)) \quad (7)$$

Here, $L(\theta)$ is the average loss over N training samples, X_i and y_i are the input and label for the i^{th} sample, f is the model with parameters θ initialised from a pretrained model, and ℓ is the loss function, such as cross-entropy loss for classification. In practical applications, such a hybrid approach enhances the efficiency and accuracy of spatiotemporal models in practical applications [56,57].

Key considerations from a project methodology standpoint include the availability and balance of datasets, the computational limitations of intended target platforms, and the integration of traditional ML techniques for enhanced data insights. The adopted methodologies encompass transfer learning for enhanced object detection and classification, leveraging deep learning for feature generation, and utilising traditional ML classifiers to extract further insights, ensuring a comprehensive approach to developing ARDAD solutions.

3. Materials and Methods

This paper documents a critical phase in the development process, transitioning from the initial Proof-of-Concept (PoC) [7] to a minimum viable product (MVP) for the Auckland Harbour Bridge’s movable concrete barrier (MCB), addressing additional research problems related to detecting anomalies in a noisy spatiotemporal scenario.

In addressing real-world challenges and producing technological solutions, the research methodology had to be significantly adapted to adhere to NZTA Waka Kotahi’s safety rules associated with the recording of videos showing pins deliberately set in tem-

porary unsafe positions (on the short barrier segments located on the safe working area before the Harbour Bridge). The data collection equipment is safely attached to a BTM using makeshift contraptions (Figure 4 and Table 1). However, leveraging multiple Barrier Transfer Machine (BTM) operations to collect sparse minority-class data was not viable, leading to the development of a background cloning approach to produce synthetic frames depicting unsafe pin positions (Table 2). Such a methodological approach is aligned with multidimensional problem solving for projects where ongoing adaptation to external factors is necessary [58,59].



Figure 4. Data collection setup: iPhone 13 Pro (for LiDAR), Samsung A7 Mobile, Apple iPad 6, External power bank, GoPro cameras and mounting equipment, and camera and iPhone mounting on a BTM.

Table 1. Data collection overview. Equipment, camera specifications, weather conditions, challenges encountered, and technological advancement progression across three data collection sessions.

Session No.	Equipment	Camera Specs	Weather Conditions	Challenges	Tech. Integration
1st	GoPro 8, GoPro 5, Samsung A7 Mobile, Apple iPad 6, duct tape, power bank, and mounting strips on the BTM. Barrier Transfer Machine (BTM) used for maintaining optimal camera angles during recording sessions	Resolution: 720p, Frame Rate: 240 fps, Field of View: Narrow, Audio: Wind Only, Protune: Enabled, White Balance: Auto, Colour: Flat, Shutter: Auto, ISO Limit: 6400, Sharpness: High, Audio Protune: Medium, Auto-rotation: Auto	Occasional rain and overcast	Vibrations, Camera Heating, Mounting Integrity Checks	Initial session: groundwork for incremental data collection
2nd	GoPro 8 and 5 on BTM's front and rear arms, other equipment same as first	Same as first session specs	Sunny	Camera Heating, Waterproofing, Battery Autonomy at High Frame Rates	The session focused on sunny and ideal lighting conditions

Table 1. Cont.

Session No.	Equipment	Camera Specs	Weather Conditions	Challenges	Tech. Integration
3rd	GoPro 9 and iPhone 13 Pro in hard case with 3D point cloud app on BTM's front arm, GPS enabled on both devices. iPhone 13 Pro utilized with a specialized application to create 3D point cloud data, enhancing the depth and quality of spatial analysis	GoPro 9: 1080p resolution, 240 fps, Upgraded specs for harsh conditions iPhone 13 Pro: Lidar-enabled camera for 3D point cloud data capture and GPS for location tracking	Early morning dark and rain	Ensuring Camera Waterproofing, Maintaining Battery Autonomy, Device Stability in Harsh Conditions	Integration of Lidar and GPS for advanced spatial data capture; utilised 5G for real-time data transmission. Use of 5G technology was considered to enhance real-time data capture and transmission, ensuring that large datasets could be managed effectively

3.1. Step 1: Data Collection and Synthetic Data Generation

Incremental data collection began with the first session under occasional rain and overcast conditions, setting a precedent for the resilience of the process. During the second session, conducted in sunny weather, two GoPro cameras were mounted on the arms of the Barrier Transfer Machine (BTM), with technical challenges such as vibrations, camera overheating, the need for waterproofing, and maintaining battery life at high frame rates becoming apparent. The third session was conducted in heavy rain, prompting further protocol and system design enhancements. Concurrently, integrating advanced technologies like Lidar, GPS, and 5G networks was considered to augment the robustness and capability of the data collection cycles (Table 1).

The pre-emptive manual inspections of moveable concrete barriers (MCBs) before any BTM operation for lane modification eliminated the possibility of naturally capturing the required unsafe pin positions (Figure 5). The precautionary measure is understandable given the substantial risk posed by potentially loose concrete blocks during the MCB transfer process, which could compromise the BTM's integrity and the safety of road users.

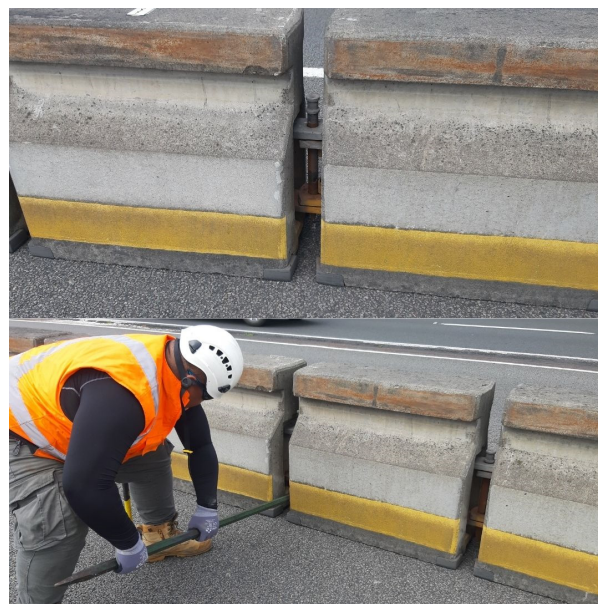


Figure 5. Pin manually pushed out of place by NZTA staff at the author's request. The process was challenging and labour-intensive, and the staff member needed several minutes per pin for adjustments. Note: Such an option was not viable as it does not represent the natural environment of the ROI where real-world problems need solving.

Table 2. Semi-automated process for generating a diverse dataset of synthetic pin positions (based on the manual cloning method illustrated in Figure 6).

Algorithm: Semi-Automated Synthetic Image Generation for Pin Position

Input: Series of Pin_OK images, Pin_OUT template image

Parameters:

- Alignment normalisation method
- Segmentation method
- Displacement calculation method
- Transformation method
- Background reconstruction method
- Edge refinement method

Output: Synthetic dataset with varied and accurately positioned Pin_Out frames

1: Load Images

- 1.1: Load the series of Pin_OK images
- 1.2: Load the Pin_OUT template image

2: Normalise alignment

- 2.1: Normalise the alignment of the pin in the series of Pin_OK images to a standard reference position

3: Process each normalised Pin_OK image

- 3.1: Segment the pin using a region-based segmentation method
- 3.2: Calculate the displacement needed based on the Pin_OUT template
- 3.3: Apply the calculated displacement to adjust the pin position and create a synthetic Pin_Out image
- 3.4: Reconstruct the background where pin was initially placed
- 3.5: Refine the edges of the moved pin to ensure seamless integration
- 3.6: Visually validate the quality of the synthetic image
- 3.7: If the quality is acceptable, add the synthetic image to the dataset
- 3.8: If adjustments are needed, refine the parameters and repeat the process

4: Repeat for all images

- 4.1: Continue the process for all Pin_OK images in the series to create a comprehensive dataset with varied Pin_Out positions

Faced with the sparsity of unsafe pin position, a novel synthetic data creation method involving background cloning from original video frames was introduced (Figure 6). Cloning allows for generating varied representations of the minority class, thus addressing the skewed data distribution. Initial attempts yielded imperfect frames with jagged edges, but the process was honed through iterative refinement, resulting in high-fidelity synthetic frames that significantly bolstered the dataset. Adding the synthetic frames facilitated the fine-tuning of model precision and recall, which is crucial for the efficacy of automated inspection systems and ensures higher sensitivity towards false negatives—a priority for traffic safety systems. The cloning approach proved pivotal in circumventing the limitations imposed by manual data collection methods, paving the way for a safer, more efficient means of training robust detection models.

As shown in Figure 6, creating synthetic frames using an image editor takes at least 20 min per frame, highlighting the need for a more efficient and automated solution (Table 2). The process begins by loading a series of images where the pin is in a safe position (Pin_OK) and a template image where the pin is in an unsafe position (Pin_OUT). The method involves several key steps: normalising the alignment of the pins in each Pin_OK image to a standard reference position, segmenting the pins using a region-based segmentation method, and calculating the displacement needed to replicate the Pin_OUT template.

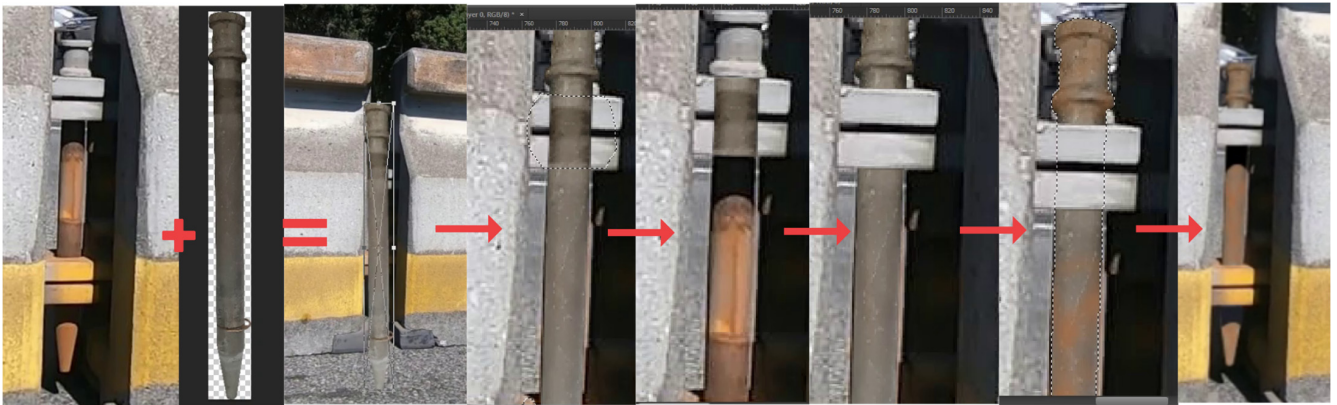


Figure 6. Synthetic data generation: The image illustrates the various stages of manipulating a video frame to create synthetic images depicting a metal pin in unsafe positions. Starting with a standalone image of the metal pin, the process involves adjusting its orientation, position, and environmental context to generate realistic, unsafe scenarios.

3.2. Step 2: Data Augmentation

Data augmentation was necessary to work with sparse and unbalanced datasets. We also considered commonly used approaches to enhance the training robustness of neural networks by generating additional training data from existing datasets and helping to prevent overfitting [60]. The augmentation techniques applied included geometric and affine transformations where rotation, resizing, reflection, translation, and shearing were utilised to modify the image structure without altering its content [37] (Figure 7). The transformations were implemented using MATLAB's Image Processing Toolbox, which facilitated the efficient application of such techniques. Additionally, this research incorporated methods to introduce realistic variations into the dataset by adding noise and blur effects, specifically using Gaussian and Salt and Pepper noise patterns. Such effects are applied using MATLAB's `imnoise` function for noise addition and `imgaussfilt` for Gaussian blur. Such modifications simulated potential real-world imperfections in data, aiding the network in learning to handle such irregularities effectively.

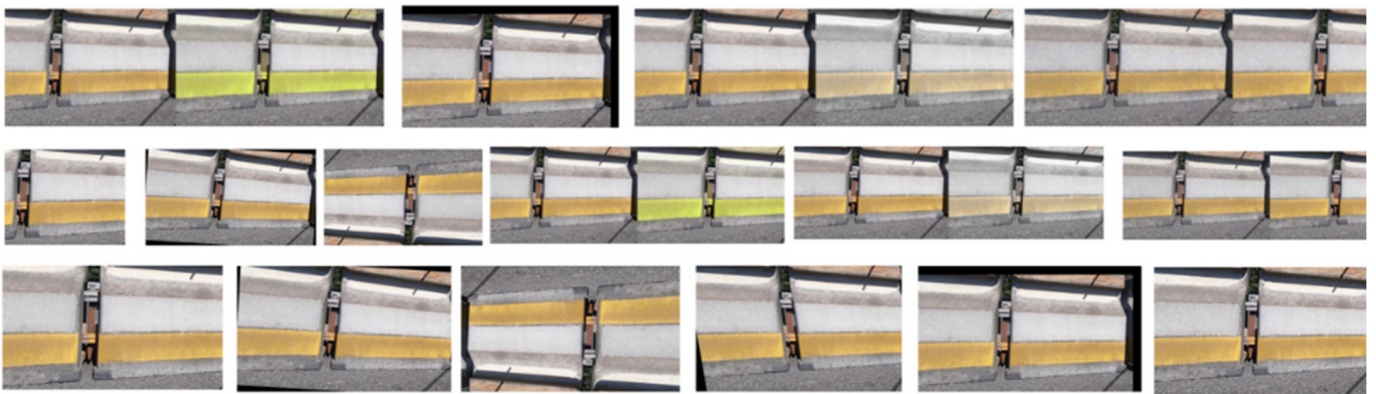


Figure 7. Illustrative outputs from data augmentation techniques showcase the range of image transformations applied to enhance model training. It includes geometric transformations such as rotation and scaling, affine transformations with various translations and shearing, and visual effects such as Gaussian blur, noise addition, and sophisticated colour adjustments. The modifications are instrumental in preparing the neural network to handle diverse and realistic scenarios encountered in practical applications, Xu, Yoon [61].

Further innovation includes the adoption of advanced colour transformation techniques. Using MATLAB's `jitterColorHSV` function, the images underwent random adjust-

ments in brightness, contrast, hue, and saturation, broadening the range of visual data the model was exposed to during training. Additionally, a novel colour transformation approach that combines table lookup methods and 3D colour space interpolation allowed for high-quality, real-time colour processing.

The comprehensive approach to data augmentation diversifies the training set, enhancing the model’s generalisation capabilities, which are important for similar applications, including smart city contexts and the application of technology to improve the usability of spaces where human activity may occur.

3.3. Step 3: Data Distribution Analysis Post Minority Boosting

Synthetic frames balanced the dataset, enhancing classification accuracy on a relatively small, manually labelled dataset (Figure 8). This approach was advantageous under restrictive conditions restricting minority class data collection. Additionally, synthetic data facilitated fine-tuning model performance metrics such as Precision and Recall, focusing on minimising false negatives due to their criticality for safety.

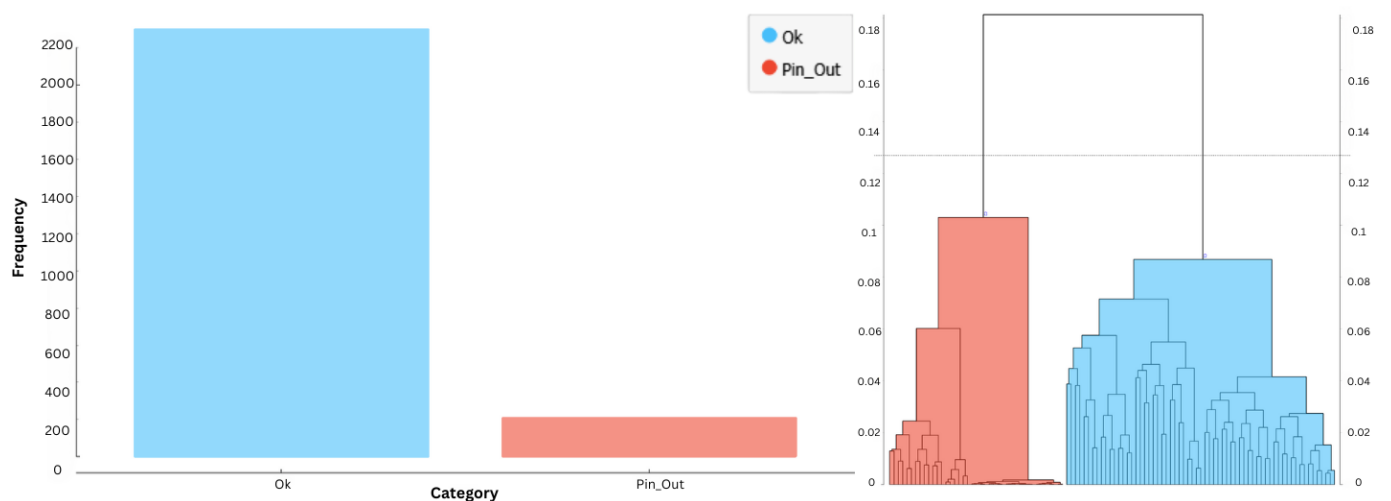


Figure 8. The dendrogram graph, derived from the data shown by the bar graph, shows two clusters (Pin_OK (red) and Pin_Out (blue)) and a visual separation of the generated multidimensional feature space.

From approximately 30,000 frames, 2300 images showing safe pin positions were curated. To address the challenge of limited occurrences of unsafe pin conditions—a common issue in training robust detection systems—210 synthetic images simulating unsafe pin positions were generated and added to the dataset (Table 3).

Table 3. Data distribution: overview of the number of Pin_OK and Pin_Out images used for training and validation across different video recording sessions.

Training Video Session No. (Table 1)	Training Data		Validation Data	
	Pin_OK Images	Pin_Out Images	Pin_OK Images	Pin_Out Images
1	155	32	40	5
2	725	80	180	20
2	800	40	200	10
3	400	16	100	5

Creating varying degrees of ‘Pin_Out’ positions helped optimise the balance between false positives and negatives, which is crucial for refining the model’s accuracy and robustness (Table 4).

Table 4. Initial classification results achieved from data clusters from Figure 8. Adopted from Bačić, Rathee [7].

Confusion Table				
Model		Actual		
			PIN_OK	PIN_OUT
Logistic regression	P r e d i c t e d	PIN_OK	[98.5%	1.5%
		PIN_OUT	0	100%]
Neural Network		PIN_OK	[98.9%	1.1%
		PIN_OUT	0	100%]
SVM		PIN_OK	[98.1%	1.9%
		PIN_OUT	0	100%]

Synthetic frames varying in Pin_Out positions improved the tuning of false positive (FP) and false negative (FN) ratios. The aim is to achieve $FP > FN$ using a dataset with borderline unsafe pin positions, enhancing classification accuracy. Accuracy was measured using the following formula:

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \quad (8)$$

TP represents the true positives or correctly identified Pin_Out frames, and TN represents the true negatives or correctly identified Pin_OK frames. Precision, the ratio of correctly predicted positive observations to the total predicted positives, is defined as follows:

$$\text{Accuracy} = \frac{TP}{TP + FP} \quad (9)$$

The comprehensive approach involved a meticulous three-fold stratified cross-validation process, as detailed in Table 5.

Table 5. The cross-validation test and score results updated from initial research.

3-Fold Stratified Cross-Validation		
Model	Precision	Recall
Logistic regression parameters: Regularisation: Ridge (L2), C = 1	0.995	0.995
Multilayer Perceptron (MLP)- Parameters: Hidden layers: 2 Neurons: Activation function: ReLu Solver: Adam Alpha: 0.02 Max. iterations: 200 Backpropagation algorithm	0.995	0.995
Support Vector Machine (SVM) parameters: C = 1.0, $\epsilon = 0.1$ Kernel: Linear Numerical tolerance: 0.001 Max. iteration: 100	0.985	0.984

3.4. Step 4: Detecting Region of Interest (ROI)

Our research team comprehensively investigated various Region of Interest (ROI) detection methodologies in developing an automated pin detection system for traffic management applications. Each method was scrutinised for its ability to accurately identify pin locations within video frames under various environmental conditions. Shuffling through different methodologies illuminated each approach's challenges and potential, culminating in us finding a particularly effective solution. Initially, our exploration utilised region proposal methods, which calculate bounding boxes around potential ROIs (Figure 9C). While initially promising, the approach necessitated frequent manual adjustments to ensure precision in detection and counting, rendering it less feasible for dynamic real-world applications where automation is paramount. A Gaussian Mixture Model (GMM) was employed for adaptive background modelling (Figure 9B). The GMM was adept at segmenting moving pins from static backgrounds by modelling the probability density distribution of pixel intensities. Despite its effectiveness in foreground differentiation, the GMM required extensive post-processing to localise each pin accurately, limiting its standalone utility in precise ROI detection. We also explored colour-based segmentation using K-Means clustering within RGB, LAB, and HSV colour spaces—the latter of which is particularly favoured due to its resilience against variations in lighting (Figure 9A). The method exploited the unique colour signatures of the pins but faced limitations in specificity, often capturing unrelated objects sharing similar colour profiles.

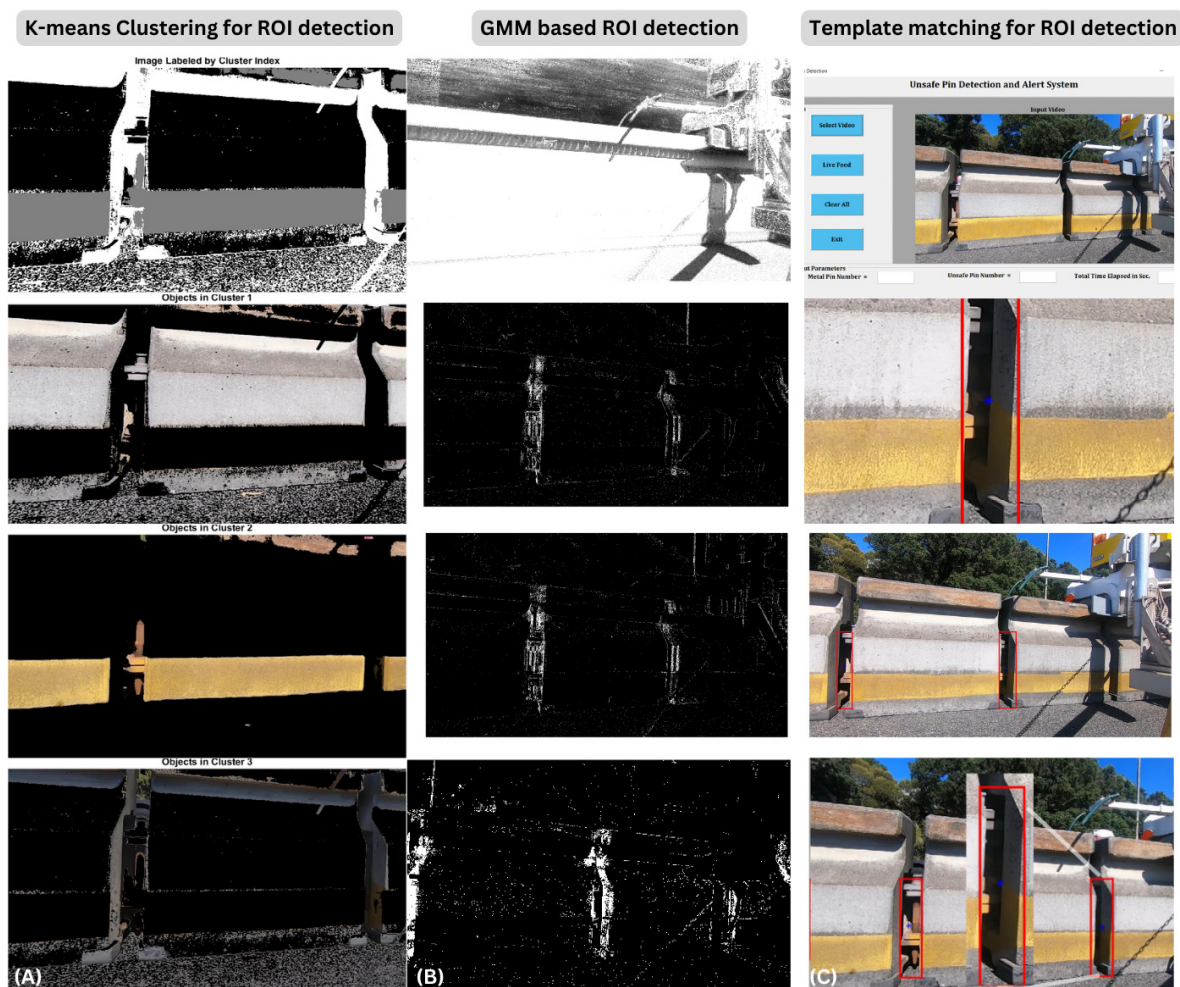


Figure 9. Comparative analysis of promising ROI detection techniques in automated pin detection: (1) colour-based segmentation using K-Means clustering in HSV colour space, highlighting the unique

colour signatures of pins while addressing challenges in specificity. (2) Gaussian Mixture Model (GMM)-based detection illustrates foreground–background segmentation and pin movement tracking between concrete blocks. (3) Regionprops application, showcasing template matching and automated bounding box detection for precise localisation and reduced manual intervention in pin ROI detection and labelling.

The methods tried in Figure 9 led to us adopting MATLAB's *regionprops* function, which proved superior in addressing previous methods' precision and automation challenges. The *regionprops* analyses each connected a component in binary video frames, effectively quantifying the area and bounding box coordinates for each detected ROI. The function enhanced the accuracy of pin detection and facilitated a significant reduction in manual intervention by automating the ROI detection process.

The underlying mathematical concepts of the *regionprops* function are integral to its performance (Equations (8)–(13)).

Area: The area of a region is computed as the number of pixels within it.

$$\text{Area} = \sum_{(i,j) \in R} 1 \quad (10)$$

In this equation, the area represents the total number of pixels in the region R. Each pixel within the region contributes a value of 1 to the total area, effectively counting the pixels.

Centroid: The centroid is determined by averaging the positions of all the pixels in the region, as follows:

$$\text{Centroid}_x = \frac{1}{N} \sum_{(i,j) \in R} i, \text{Centroid}_y = \frac{1}{N} \sum_{(i,j) \in R} j \quad (11)$$

Bounding box: The bounding box, which encloses the region, is defined by the minimum and maximum x and y coordinates of the pixels.

$$\text{BoundingBox} = [x_{\min}, y_{\min}, x_{\max} - x_{\min} + 1, y_{\max} - y_{\min} + 1] \quad (12)$$

Major and minor axis length: For regions approximated by ellipses, the major and minor axis lengths are derived from the eigenvalues of the covariance matrix of the pixel coordinates.

$$\text{MajorAxisLength} = 2\sqrt{\frac{\lambda_1}{N}}, \text{MinorAxisLength} = 2\sqrt{\frac{\lambda_2}{N}} \quad (13)$$

Orientation: The orientation of such an ellipse, indicating the angle between the x-axis and the major axis, is calculated using the second moments of the region.

$$\text{Orientation} = \frac{1}{2} \arctan\left(\frac{2\mu_{11}}{\mu_{20} - \mu_{02}}\right) \quad (14)$$

Perimeter: The perimeter is measured by summing the distances between consecutive boundary pixels.

$$\text{Perimeter} = \sum_{\text{boundarypixels}} \text{distance between consecutive boundary pixels} \quad (15)$$

The integration of *regionprops* into our method significantly advances the automation of pin detection. The enhancement offers high precision and efficiency, which are critical for traffic management systems requiring reliability and rapid processing.

Our successful implementation emphasises the importance of a methodical approach in engineering research, demonstrating how iterative testing and the integration of various techniques can tackle complex real-world problems. This study highlights the effectiveness of combining traditional image processing methods with advanced analytical tools and sets a new standard for future research in automated traffic safety and management systems.

The findings provide a robust framework for researchers and practitioners to enhance automated detection systems across engineering domains.

3.5. Step 5: Model Training and Validation

The computing system (Table 6) provided sufficient hardware capabilities to train the model at the core of the developed 'Unsafe Pin Detection and Alert System' system.

Table 6. A development system using NVIDIA GPU parallel processing architecture.

System Configuration	
Processor	Intel Core i7 Processor
Memory	32 GB RAM
Hard Drive	512 GB Solid State Drive
Graphics	NVIDIA GeForce RTX2070 Super 8 GB GFX
Operating System	Windows 10

ResNet-50, enhanced with 3D convolutions within the STENet framework, was selected after an extensive comparison with other models like Squeezenet, GoogLeNet, InceptionV3, and Mobilenetv2. ResNet-50 is favoured due to its optimal balance between computational efficiency and performance, confirmed by satisfactory GPU RAM evaluations during initial experiments (Table 7).

Table 7. Comparison of pretrained deep neural networks based on input image resolution, number of parameters, depth, and model size.

Pretrained Deep Neural Networks				
Model	Input Image Resolution	Parameters (1,000,000)	Depth	Size
AlexNet	227 × 227	61	8	227
SqueezeNet	227 × 227	1.24	18	4.6
GoogleNet	224 × 224	7	22	27
Inception v3	299 × 299	23.9	23.9	48
MobileNet v2	224 × 224	3.5	3.6	53
Resnet 50	224 × 224	25.6	50	96

Various environmental elements, such as rust-coloured foliage, closely matched the appearance of rusted metal pins, making visual differentiation challenging. Similarly, unrusted metal pins shared the same colour as the road surface and the tyres of passing vehicles, further complicating the identification process. The metal barriers in the background often had shapes resembling the horizontal profile of the metal pins, while the dynamic colours of passing vehicles introduced intense background variations. In addition to labelling the pin statuses, regions of interest (ROIs) were defined on each image using the *regionprops* method to distinguish between relevant features and potential background noise. This method, which analyses each connected component in binary video frames to quantify the area and bounding box coordinates for each detected ROI, significantly enhanced the accuracy of pin detection and reduced manual intervention. The combined consideration of shape and colour homogeneity was a significant factor in training the AHB pin detection model (Figure 10).

Additionally, certain parts of the movable concrete barrier (MCB) system closely resembled the shape of the metal pins, creating further confusion (Figure 11(3)). These factors made the detection task challenging, as the system had to differentiate between multiple visually similar elements under varying conditions. The dataset then underwent a series of image augmentation processes to simulate various operational conditions not covered by the initial video capture. The processes included geometric transformations

such as rotations, scaling, translations, and adding image perturbations like Gaussian blur and noise, which are crucial for training models to perform well under practical deployment conditions.

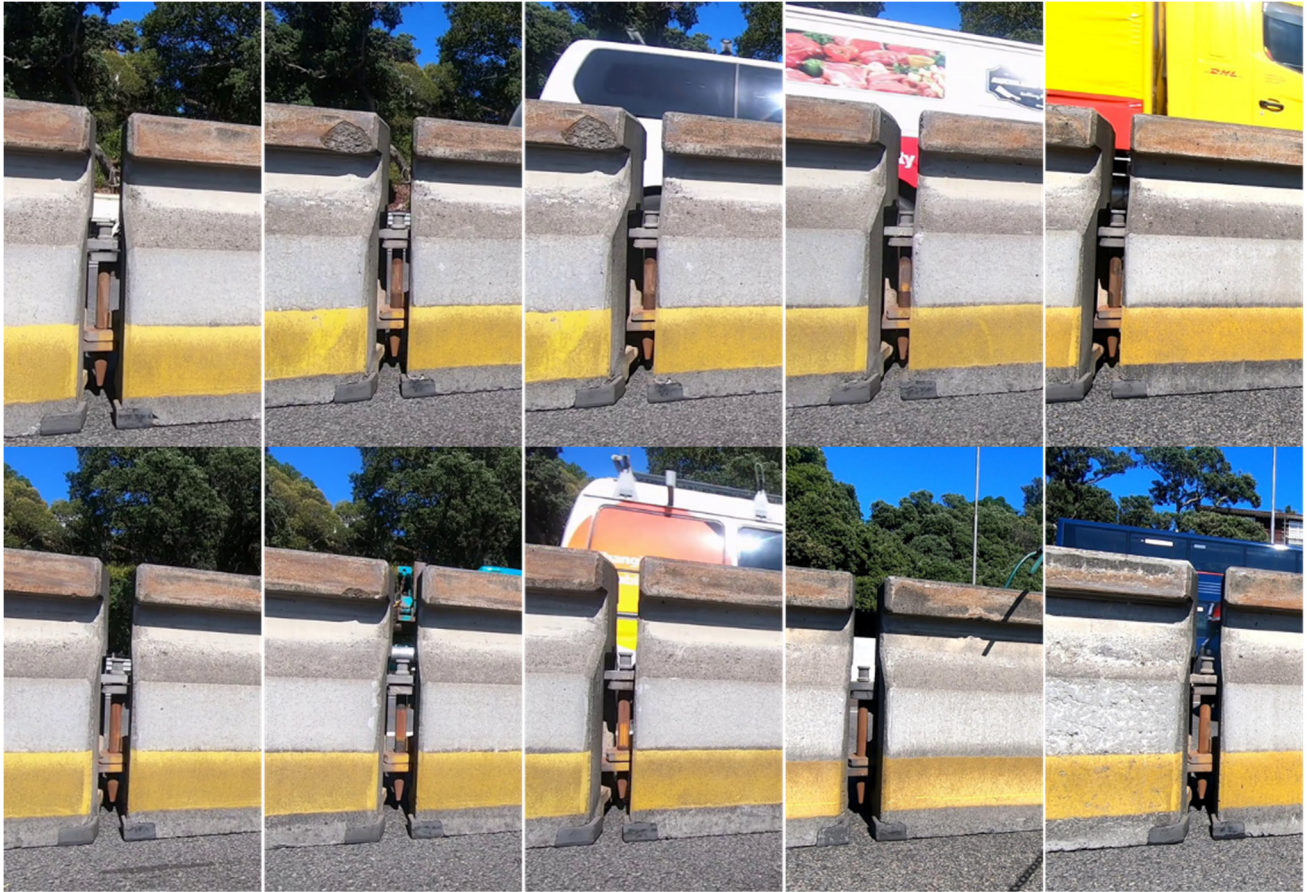


Figure 10. Illustration of the background challenges in ROI detection.

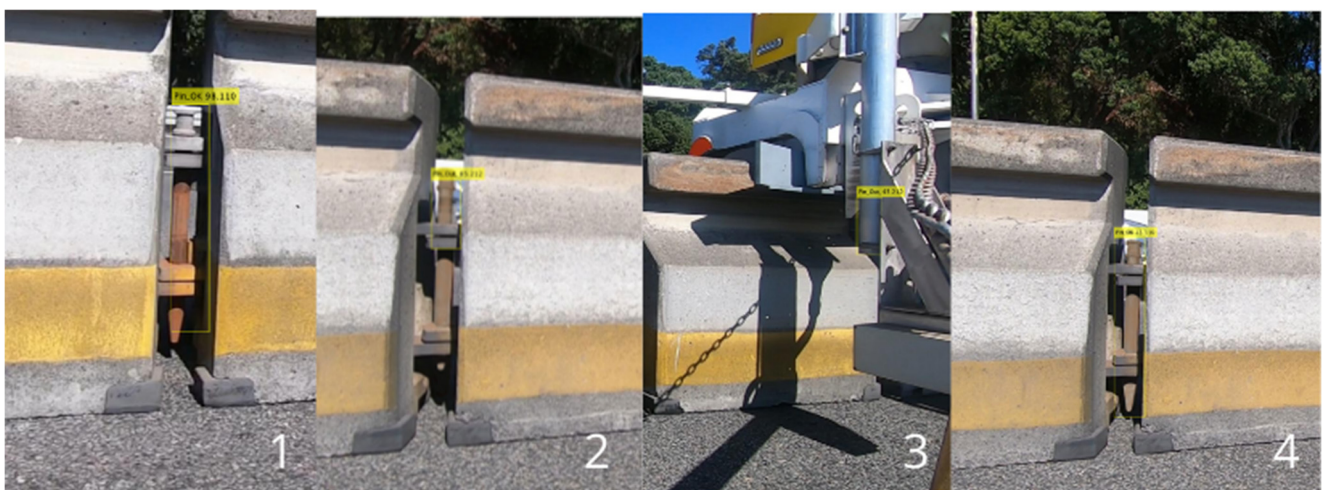


Figure 11. Pin detection examples mixed with false-positive detection (3) illustrate a high level of background noise that looks similar to features of the ROI.

A critical component of the training process was the adjustment of the learning rate, a parameter that significantly impacts the convergence and final performance of the model.

Learning rates ranging from 0.01 to 0.00001 using ResNet-50 were tested (Figure 12). The optimal rate was determined through a series of trials evaluating model accuracy and loss metrics. The experiments utilised ROC curves to visually represent the trade-offs between true-positive and false-positive rates at various threshold settings, enabling an informed selection of the best-performing model under the given training conditions. The results consistently showed that a learning rate of 0.0001 provided the best balance between training speed and model accuracy.

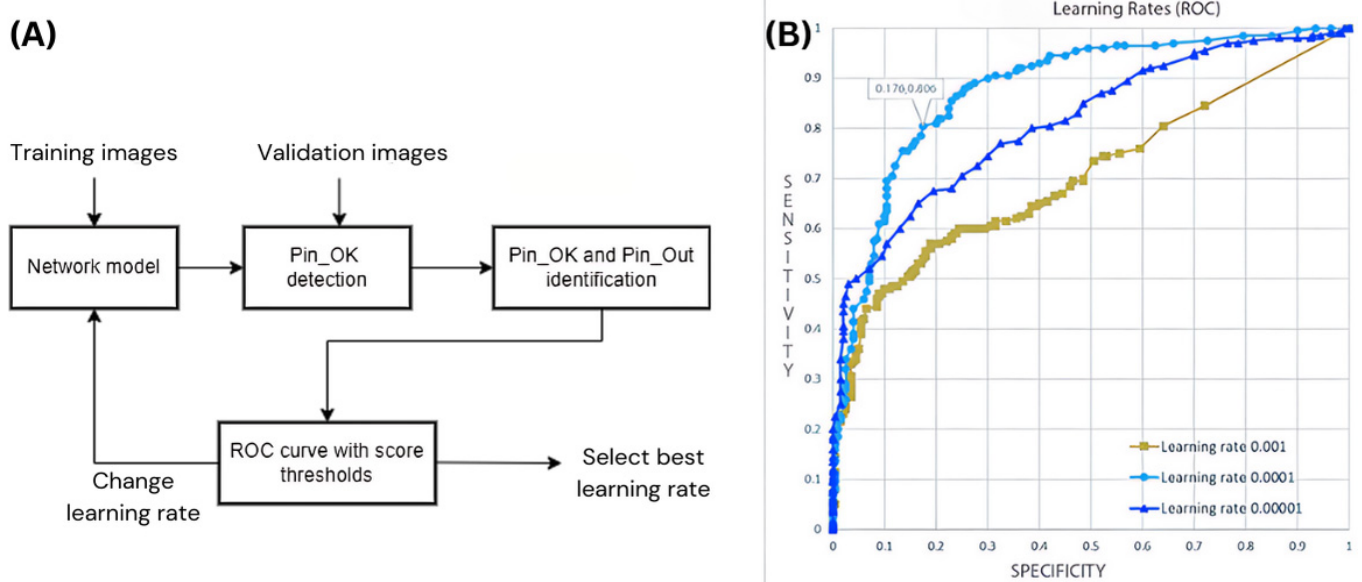


Figure 12. The training process to find the best network model based on learning rate: (A) flow of the training process and (B) ROC curves for different learning rates showing the performance of learning rates. The learning rate of 0.0001 was chosen as it balances training speed and accuracy, providing better performance while avoiding the high computational cost and time.

In conclusion, this research detailed the selection, preparation, and training of a deep learning model capable of detecting and classifying metal pin statuses in movable concrete barriers. The methodological approach, emphasising the creation of a comprehensive and diverse training dataset through real and synthetic images and rigorous model training protocols, demonstrates a scalable and efficient solution to enhance public infrastructure safety through advanced AI techniques.

3.6. Step 6: Model Analysis and Testing

The STENet architecture, inspired by and building upon the robust framework of ResNet-50, exhibits considerable potential for spatiotemporal anomaly detection tasks, such as those needed for monitoring the Auckland Harbour Bridge (AHB). While ResNet-50 was initially developed for image classification, its strong spatial feature extraction capabilities make it a solid foundation for further enhancements to analyse video and live feed data. STENet integrates components from ResNet-50 with 3D convolutions to capture spatial and temporal information, facilitating the direct analysis of motion and dynamic changes within video sequences. Further enhancements involve leveraging ResNet-50's spatial feature extraction capabilities and Recurrent Neural Networks (RNNs) to track temporal sequences or employing temporal pooling methods to summarise video segments efficiently.

By adopting strategies for model simplification, edge computing, and incremental learning, STENet enhances system efficiency and responsiveness to new anomalies. This makes STENet particularly suited for continuous surveillance and traffic monitoring applications. Table 8 highlights the performance metrics of STENet compared to other models. By incorporating adaptations of ResNet-50's core features, STENet effectively recognises

appearance and behavioural pattern deviations over time, providing a robust framework for real-time anomaly detection across diverse operational environments (Figure 13).

Table 8. The Spatio-Temporal Enhanced Network (STENet) achieves an outstanding 95.2% accuracy score and an F1-Score of 94.8%, showcasing its exceptional capability to navigate through the challenges of background noise and small ROIs, with a remarkable ROC-AUC of 98.5%, solidifying its robustness in class differentiation in spatio-temporal tasks.

Model	Accuracy	Precision	Recall	F1-Score	ROC-AUC	Training Time	Inference Time	Number of Parameters
STENet	95.2%	94.5%	95.2%	94.8%	98.5%	4 h	80 ms	25 M
VGG-19	88.0%	88.5%	89.0%	88.7%	94.0%	6.5 h	90 ms	142 M
ResNet-50	90.0%	90.3%	91.0%	90.6%	95.8%	4 h	70 ms	25.6 M
InceptionV3	91.0%	91.2%	91.5%	91.3%	96.0%	3.5 h	65 ms	23.8 M
SqueezeNet	85.0%	85.5%	86.0%	85.7%	92.0%	2.5 h	35 ms	1.25 M

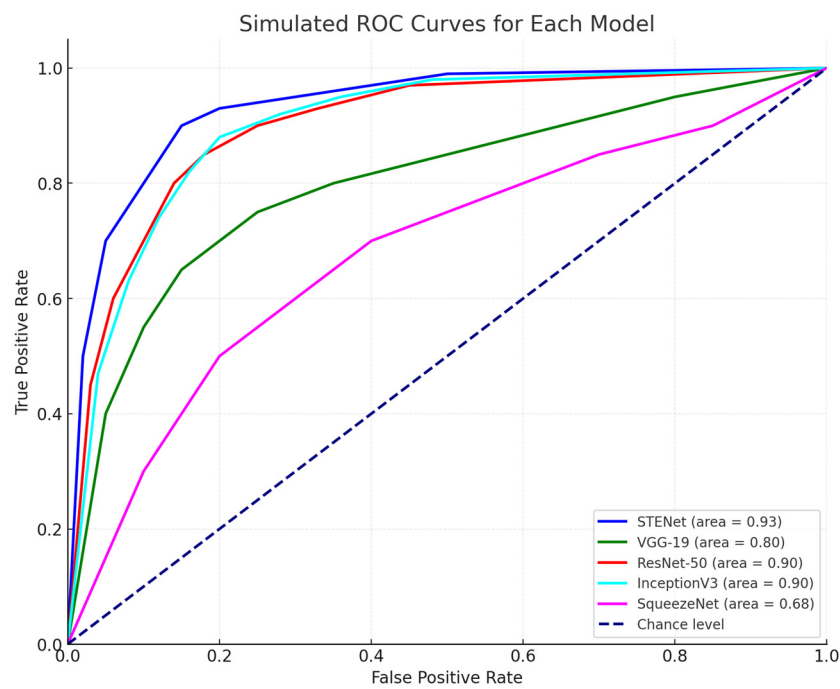


Figure 13. ROC curve comparison of Table 8.

4. Results and Discussion

The efficacy of the pin detection and alert system, which encompasses both pin Region of Interest (ROI) detection and tracking alongside pin status detection and alert functionalities, was critically evaluated. The assessments drew on methods previously detailed, with particular attention to the operational conditions influencing system performance.

4.1. Preparations, Data Recording Protocol, Findings, and Insights

The system’s performance was assessed by examining recorded videos frame by frame, focusing on pin ROI tracking. Several environmental and operational factors were identified that could adversely affect the results.

- **Lighting Conditions:** Overcast conditions significantly hampered visibility, challenging the detection of pins between concrete blocks and rendering the detection and counting processes unreliable.

- **Background Movements:** Moving vehicles introduced significant noise into the background, severely hindering detection accuracy. While the greyish hue of the road and concrete blocks helped simplify the background, the large vehicles passing by disrupted visual clarity. Additionally, the rust colours and yellow of the movable concrete barriers (MCBs) and fragments of broken concrete barriers and metal barriers often blended into the pin regions of interest (ROIs), further decreasing tracking accuracy.
- **Shadows and Lighting:** Shadows cast on pin ROIs were sometimes misinterpreted as moving objects. Darker shadows could compromise the detection and tracking systems, particularly under bright sunlight.
- **Vibrational Distortions:** Operational speeds exceeding 6 km/h induced excessive vibration, blurring the images and introducing background noise, further complicating the detection processes.

To mitigate weather-related variabilities, multiple videos recorded in sunny conditions were analysed. The first video, capturing a broader field from the front arm of the Barrier Transfer Machine (BTM), displayed approximately two ROIs per frame, shot during a clear afternoon. The second video, taken from the BTM's rear arm, showed a narrower field of view with sometimes only partial visibility of a single-pin ROI.

4.2. Creating Synthetic Frames

The concept of creating synthetic frames encountered significant challenges. The initial cloning techniques were insufficient, yielding a non-viable minority dataset, which necessitated a shift from classification techniques to traditional mathematical approaches. Subsequent consultations with graphic experts led to the adoption of advanced methods using Photoshop and Gimp for generating synthetic frames (Figure 14).

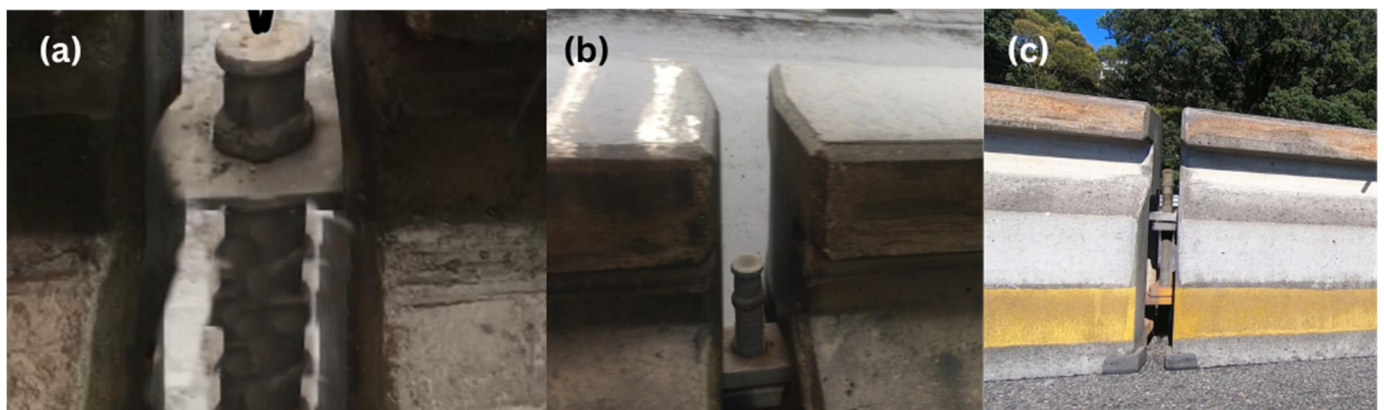


Figure 14. Cloning process to create synthetic frames depicting the 'PIN OUT' position. (a) Failed attempts until success (b,c); two distinct angles: (b) camera mounted at the back arm of the Barrier Transfer Machine (BTM) and (c) camera mounted at the front arm of the BTM for enhanced modelling accuracy.

Approximately 40 synthetic frames were crafted from video footage captured on a mobile phone during the Proof-Of-Concept phase. A second batch of 200+ synthetic frames was later produced from higher-quality video recordings, demonstrating a noticeable improvement in frame clarity between the initial and later productions. Given the labour-intensive nature of manual frame creation, efforts were made to automate the process using Photoshop's action panel. However, the complexity of the background elements hindered full automation. While MATLAB offers pixel cloning techniques, further research is required to develop a fully automated and robust method for synthetic frame generation. The analysis underscores the critical dependencies of environmental conditions on the performance of pin detection systems and highlights the ongoing need for

technical enhancements in synthetic data generation to support robust system training and validation.

4.3. System Prototype: Pin Tracking, Counting, and Alerting Functionality

The development of the metal pin detection and alert system was guided by the requirement that the end-user, presumed non-technical, has a simplified interface for efficient system operation. The system prototype, named 'Unsafe Pin Detection and Alert System' (Figure 15), was designed and implemented using MATLAB App Designer [62], integrating back-end operations with a graphical user interface (GUI).

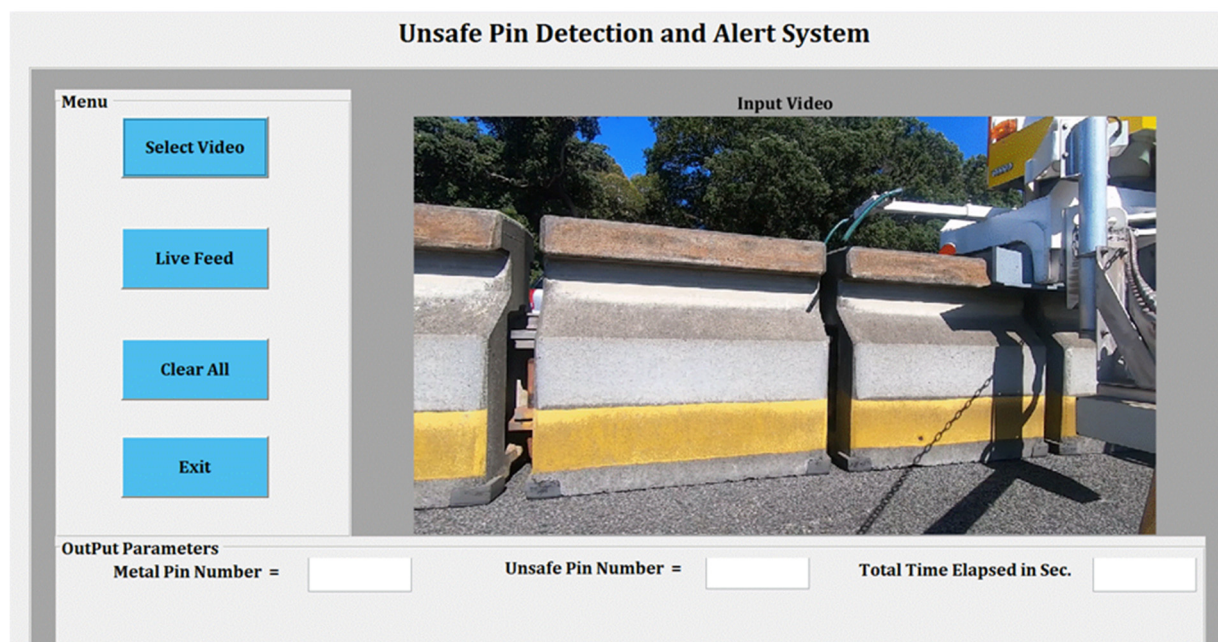


Figure 15. Graphical user interface for 'Unsafe Pin Detection and Alert System'.

System Functionalities

The App allows the end-user to upload a video or frame sequence for analysis. Upon upload, the App leverages a deep learning-based pin status detection network to analyse the video feed. The system maintains a count of metal pins, and if it detects a pin in an unsafe position, it triggers an alert and displays the specific metal pin ROI on the screen. The second functionality supports live feed analysis, where the user can connect the App to a live camera feed. The pin detection network activates to analyse incoming frames in real-time as the live feed is transmitted to the Unsafe Pin Detection and Alert System monitoring App (Figure 15). Like the pre-recorded analysis, the system tracks and counts the metal pin ROIs, alerting the user and displaying the ROI number when an unsafe pin is detected. The metal pin detection and alert system interface, as shown in Figure 15, is designed to be intuitive and easily navigable, ensuring that users can operate the system without prior technical knowledge. The App's design and operational logic were specifically tailored to meet the needs of NZTA, allowing for custom modifications to accommodate specific operational requirements or updates.

In conclusion, in theory, the alert system App is ready to be deployed. There is potential for further development into a complete product deployment, pending interest and additional support from the NZTA. Future work will focus on refining the App's functionality, enhancing its performance, and ensuring its onsite installation and integration into existing infrastructure systems. Such enhancements, however, depend on the availability of further resources and continued interest from stakeholders.

4.4. Results

The evaluation of the performance of the STENet model was aimed at effectively finding and localising the Pin_Out status using the workstation specifications outlined previously in Table 6. We applied the models to the pin's region of interest (ROI) frame-by-frame tracking and real-time counting, assigning each pin ROI an index to maintain continuity across frames. The detector model configurations are summarised in Table 9, which details parameters such as batch size, number of epochs, learning rate, and optimiser choices. Two optimisation techniques were utilised, Stochastic Gradient Descent with Momentum (SGDM) and Adam, both chosen for their ability to enhance model performance through efficient direction finding in the optimisation landscape [63,64].

Table 9. Configuration of the detector model. The table lists the batch size, number of epochs, learning rate, and optimiser used for training the detector model. Two different optimisation techniques, Stochastic Gradient Descent with Momentum (SGDM) and Adaptive Moment Estimation (Adam), were employed to enhance model performance.

Model	Batch Size	Epoch	Learning Rate	Optimizer
Detector	64	100	0.0001	Sgdm
	64	100	0.0001	Adam

Pin ROI tracking was evaluated for accuracy by comparing the consistency of index assignment from frame to frame. Discrepancies in index continuity were noted as errors in tracking. The dataset was divided into 70% training frames, 10% validation, and 20% testing frames to assess the model robustly across varied conditions. The training process using SGDM achieved quicker training times than Adam, as shown in Table 10. The validation of the models demonstrated lower Root Mean Square Error (RMSE) and validation losses with SGDM, indicating a more efficient optimisation path.

Table 10. SGDM training and validation process.

Epoch	Iteration	Time Elapsed	Mini-Batch (RMSE)	Validation (RMSE)	Minibatch Loss	Validation
10	150	00.15.05	0.91	0.87	0.8285	0.7910
20	300	00.30.20	0.74	0.83	0.5384	0.6761
30	450	00.45.12	0.66	0.78	0.4141	0.6329
40	600	01.00.15	0.62	0.76	0.3968	0.5809
50	750	01.14.50	0.56	0.77	0.3182	0.4150
60	900	01.28.58	0.53	0.76	0.2708	0.4060
70	1050	01.44.55	0.52	0.74	0.2408	0.3716
80	1200	01.58.21	0.50	0.76	0.2018	0.3208
90	1350	02.13.43	0.48	0.75	0.1808	0.2710
100	1500	02.27.23	0.47	0.74	0.1280	0.2219

The results of the pin ROI classification and the accuracy of bounding box detection by the trained STENet are depicted in Figure 16. The system was highly effective in recognising both Pin_OK and Pin_Out statuses, even when pin ROIs were partially obscured or closely positioned, which traditionally challenges detection accuracy. However, the accuracy diminished in frames where pins were too closely spaced or partially out of the field of view.



Figure 16. The selected pin ROI video frames showing generated overlays with bounding boxes.

While the system demonstrated robust performance in ideal viewing conditions, the detection accuracy varied under different fields of view and lighting conditions, emphasising the need for a more diverse training dataset. The creation of synthetic frames to augment the dataset was explored, but automation of the process remains a challenge for future work, as manual frame creation proved time-intensive. Table 11 showcases the classification metrics—accuracy, precision, and recall—after training, highlighting the model’s strong performance overall.

Table 11. Performance of the STENet where ResNet 50 is used as a classifier.

Classes	Accuracy		Precision		Recall	
	Training	Validation	Training	Validation	Train	Validation
Pin_Ok	0.952	0.945	0.942	0.940	0.940	0.945
Pin_Out	0.925	0.922	0.910	0.900	0.900	0.890

Despite variations due to the limited diversity in the minority class data, the classifier maintains high accuracy and precision across both classes. The consistent performance metrics for the ‘Pin_Ok’ and ‘Pin_Out’ classes demonstrate the model’s robustness and reliability in identifying and classifying majority and minority class instances.

The results underline the practical application of deep learning models to metal pin detection tasks, highlighting the necessity for further improvements in model training and

synthetic data generation to handle diverse operational scenarios effectively. The research's main contributions to ARDAD are listed as follows:

- Spatiotemporal analysis for automated monitoring of traffic barriers on the Auckland Harbour Bridge and other traffic locations using the same barrier to control traffic flow;
- Transforming a PoC [7] into an MVP with deployable AI algorithms for real-time ARDAD, exemplifying the translation of research into practice;
- Semi-automated synthetic data generation methods to enhance machine learning models for complex ARDAD tasks, addressing critical traffic events' data sparsity and rarity;
- Integrating machine learning with kernel manipulation for dynamic anomaly detection to improve the precision of current ARDAD systems, increasing the average detection accuracy from 0.826 to 0.939;
- Engaging in interdisciplinary collaboration to align ARDAD advancements with stakeholder requirements, merging computational research with traffic management solutions.

4.5. Discussion

The initial attempts to record high-speed videos from public transportation and personal cars presented significant challenges. Traffic flow predictions indicated that speeds could occasionally drop below 20 Km/hr on the Auckland Harbour Bridge, ideally allowing for capturing frames at 240 fps showing perfect pin alignments. However, these conditions were rarely met, and reliance on traffic jams during rush hours did not yield the desired outcomes due to erratic stoppage times and limitations of the recording equipment. Given the critical safety requirements on the Auckland Harbour Bridge, all data collection efforts were supervised by NZTA experts, who also provided access to the Barrier Transfer Machine (BTM) and the operation site, along with necessary safety briefings. To see the narrow gaps between the movable concrete segments, high-frame-rate cameras (GoPro 5, GoPro 8, and GoPro 9) were mounted on the BTM, which moved between 6 and 9 Km/hr. Recordings were made under various weather conditions and times of the day to capture diverse operational scenarios. The difficulty in finding and recording pins that were out of position significantly hindered the research process. After numerous unsuccessful attempts, synthetic frames were adopted as a viable solution. An interim report was provided to the NZTA, showcasing hierarchical clustering and a visual separation of feature vectors related to the minority output class using Pearson and Cosine correlation-based distance measures. Such computationally more demanding measures were selected over simpler ones like Euclidean due to the high dimensionality of features extracted from CNNs relative to the number of minority class samples, including the necessity to extract information invariant to lighting conditions, precipitation, or background colours from passing vehicles.

Classifications beyond binary (pin in or out of position) remain unexplored, such as scenarios where the pin ROI is wholly obscured or metal pins are partially out. Extending the binary classification to a multi-class system could allow future systems to detect various types of damage requiring different maintenance actions. Additional datasets capturing a broader range of anomalies, and more synthetic data would be required to support such enhancements, following the methodologies outlined in our synthetic data creation algorithm in Table 2. The average detection accuracy achieved was 0.93, which is commendable given the numerous challenges encountered during model training. Compared to other region-based detectors, our hybrid model offered higher accuracy and superior processing speed, handling 40 to 45 frames per second with up to 93.6% accuracy. The integration and depth concatenation layers enhanced the detection of smaller objects by incorporating low-level image details into the detection process, facilitated by a sequence of convolution, ReLU, and batch normalisation layers. The MATLAB app provides a robust platform for expanding research into future applications. In terms of 'dealing with the unknown' and research uncertainty, for the research community undertaking similar projects exceeding one or few years, it is worth considering additional challenges that are hard to predict.

Such considerations may include changes in industry partner staff, possible pandemic lockdowns, government funding, and policy updates, which require flexibility in project and data collection planning.

A summary of the practical aspects of this study are listed as follows:

- The technology offers a cost-effective automated solution to lane and general traffic safety, augmenting but not replacing human inspections.
- The system increases inspection frequency, enhances privacy, and enables the creation of digital records for analytical insights into traffic safety.
- Future scientific efforts will use a more extensive dataset to focus on adaptive model development and performance enhancement. Additional data visualisation and hybrid methodological approaches will be explored.
- For our industry partner, the NZTA, the project paves the way for independent software development and potential system integration into broader smart city infrastructures.
- The transition from a minimum viable product (MVP) to production systems will involve extensive testing, code optimisation, and, potentially, transitioning from MATLAB to Python to enhance computational efficiency integration capabilities and minimise the possibility of vendor locking into proprietary infrastructures and data processing outside of the national jurisdiction.
- This study lays the groundwork for future innovations in traffic management technology, positioning the NZTA to leverage the advancements in its ongoing modernisation of traffic infrastructure and smart city initiatives.

Future system advancements will consider enabling pin status tracking from various points of view, potentially including additional data collection protocols and technology, expanding the system's versatility and application scope.

5. Conclusions and Future Work

The Auckland Harbour Bridge plays a crucial role in Auckland's infrastructure, with traffic flows that are uneven but predictable, reversing in volume during morning and evening rush hours. Movable Concrete Barriers (MCBs) have proven effective in managing short-distance traffic bottlenecks; however, the bridge's susceptibility to various types of vibrations, particularly around its elevated central part, raises safety concerns, necessitating frequent pin inspections. Other health and safety concerns include sole reliance on manual inspections and the potential for human errors linked to inspection staff workplace safety, unhealthy spine ergonomic posture, and issues with protective gear.

To increase the safety and safety-monitoring frequency of the MCB, we developed a privacy-preserving automated monitoring system that is transferrable from data collected on the Auckland Harbour Bridge to similar contexts involving traffic flow regulation and safety monitoring applications [10]. A novel technique for generating synthetic frames was introduced to simulate various unsafe pin positions, aiding incremental model development and performance tuning. This research successfully demonstrated that the prototype can detect unsafe pin positions directly from live feeds and previously recorded video frames under varying lighting conditions (such as bright sunshine, heavy rain, and early morning 'soft' light conditions). The scarcity of video frames showing a Pin_Out status was addressed by introducing a method for creating synthetic images to enhance the modelling process. The system's expected overall performance for pin region detection, frame selection, and pin classification was anticipated to be above 80%, with individual models achieving up to 99% accuracy on a limited dataset, as shown in Table 5. These findings warrant further validation on a larger and more balanced future dataset. The pin status detection and alert system exhibited desirable precision and accuracy, with some performance decline attributed to the dataset's unbalanced nature, diverse lighting conditions, and camera angles and distance variations. Evidence from a smaller labelled dataset suggests that the system is a viable product that does not require further intensive

manual labelling. Integrating a hybrid model facilitated the analysis and provided flexibility for future model adjustments with minimal data labelling requirements.

Future work will include further video data collection, including additional videos recorded by the NZTA and AHB maintenance teams. The enhanced data collection is expected to bolster the foundational system and help further develop universally applicable ARDAD systems for similar traffic safety contexts globally. While the pin status detection and classification results are promising, there is significant potential for further advancements in integrating pin ROI tracking with the alert system. Future iterations of the system may also leverage advanced technologies such as LiDAR and GPS, which are becoming increasingly common in modern mobile devices. Developing additional capabilities will involve extensive system training and adaptation on enriched datasets that capture various pin conditions and scenarios, potentially leading to more robust and responsive traffic management solutions.

Author Contributions: Conceptualisation, B.B. and M.R.; Software, M.R.; Validation, M.R.; Data curation, M.R.; Writing—original draft, M.R.; Writing—review and editing, B.B. and M.D.; Supervision, B.B. and M.D. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Dataset available on request from the authors.

Acknowledgments: We extend our heartfelt thanks to Gary Bonser, Martin Olive, David Ranby, and Angela Potae from the NZTA and Auckland System Management (ASM) for their invaluable assistance with transport, safety briefings, video recordings, project insights, and ongoing support. We are also grateful to the Auckland University of Technology (AUT) for funding and PhD stipend support and for providing access to the computing hardware, library, and recording equipment. We also wish to acknowledge the documentation and software development efforts from industry, academia, and open-source communities producing MATLAB, Orange Data-Mining, SqueezeNet, TensorFlow, ImageNet, Google Cloud, and OpenCV.

Conflicts of Interest: The authors declare no conflict of interest.

Notations

The following table summarises the important symbols and mathematical notations used in this paper:

Symbol	Description
Y	Luminance component in the YCbCr colour space
x_c	Centroid location of an object in template matching
x_i	Pixel location in template matching
w_i	Pixel intensity in template matching
$P(X_t)$	Probability of observing a pixel value X at time t in a Gaussian Mixture Model
$w_{k,t}$	Weight of the k -th Gaussian component at time t
$\mu_{k,t}$	Mean of the k -th Gaussian at time t
$\Sigma_{k,t}$	Covariance matrix of the k -th Gaussian at time t
ΔMt	Motion detection metric combining frame differences and optical flow
F_t	Frame at time t
∇I_t	Gradient of the image at time t
v_t	Optical flow vector at time t
α	Weighting factor in motion detection equation
K_k	Kalman gain at time k
z_k	Actual measurement at time k
H_k	Measurement matrix that maps the state space into the measurement space

References

1. New Zealand Transport Agency Waka Kotahi. Auckland Harbour Bridge Factsheet. 2024. Available online: <https://www.nzta.govt.nz/assets/site-resources/content/about/docs/auckland-harbour-bridge-factsheet.pdf> (accessed on 10 April 2024).
2. Te Waihanga New Zealand Infrastructure Commission. New Zealand's Infrastructure Asset Value, Investment, and Depreciation, 1990–2022. 2023. Available online: <https://tewaihanga.govt.nz/our-work/research-insights/build-or-maintain> (accessed on 10 April 2024).
3. New Zealand Transport Agency Waka Kotahi. How to Move a Concrete Motorway Barrier. 2024. Available online: <https://www.nzta.govt.nz/media-releases/how-to-move-a-concrete-motorway-barrier/> (accessed on 10 April 2024).
4. Yang, X.; Zhang, J.; Liu, W.; Jing, J.; Zheng, H.; Xu, W. Automation in road distress detection, diagnosis and treatment. *J. Road Eng.* **2024**, *4*, 1–26. [[CrossRef](#)]
5. Bai, D.; Li, G.; Jiang, D.; Yun, J.; Tao, B.; Jiang, G.; Sun, Y.; Ju, Z. Surface defect detection methods for industrial products with imbalanced samples: A review of progress in the 2020s. *Eng. Appl. Artif. Intell.* **2024**, *130*, 107697. [[CrossRef](#)]
6. Trilles, S.; Hammad, S.S.; Iskandaryan, D. Anomaly detection based on artificial intelligence of things: A systematic literature mapping. *Internet Things* **2024**, *25*, 101063. [[CrossRef](#)]
7. Bačić, B.; Rathee, M.; Pears, R. Automating inspection of moveable lane barrier for Auckland harbour bridge traffic safety. In Proceedings of the Neural Information Processing: 27th International Conference, ICONIP 2020, Bangkok, Thailand, 23–27 November 2020; Proceedings, Part I 27. Springer: Berlin/Heidelberg, Germany, 2020; pp. 150–161.
8. Klarák, J.; Andok, R.; Malík, P.; Kuric, I.; Ritomský, M.; Klačková, I.; Tsai, H.-Y. From anomaly detection to defect classification. *Sensors* **2024**, *24*, 429. [[CrossRef](#)] [[PubMed](#)]
9. Lozano-Ramírez, N.E.; Sánchez, O.; Carrasco-Beltrán, D.; Vidal-Méndez, S.; Castañeda, K. Digitalization and sustainability in linear projects trends: A bibliometric analysis. *Sustainability* **2023**, *15*, 15962. [[CrossRef](#)]
10. Wikipedia. Barrier Transfer Machine. 2024. Available online: https://en.wikipedia.org/wiki/Barrier_transfer_machine (accessed on 20 April 2024).
11. Rathee, M. *Safety Screening of Auckland's Harbour Bridge Movable Concrete Barrier*; Auckland University of Technology: Auckland, New Zealand, 2021.
12. Baccari, S.; Hadded, M.; Ghazzai, H.; Touati, H.; Elhadef, M. Anomaly detection in connected and autonomous vehicles: A survey, analysis, and research challenges. *IEEE Access* **2024**, *12*, 19250–19276. [[CrossRef](#)]
13. Rathee, M.; Bačić, B.; Doborjeh, M. Automated road defect and anomaly detection for traffic safety: A systematic review. *Sensors* **2023**, *23*, 5656. [[CrossRef](#)] [[PubMed](#)]
14. Cui, Y.; Liu, Z.; Lian, S. A survey on unsupervised anomaly detection algorithms for industrial images. *IEEE Access* **2023**, *11*, 55297–55315. [[CrossRef](#)]
15. Cottrell, B.H. Evaluation of a Movable Concrete Barrier System. [Tech Report] 1994. Available online: <https://rosap.nrl.bts.gov/view/dot/19352> (accessed on 9 April 2024).
16. Poe, C.M. Movable concrete barrier approach to the design and operation of a contraflow HOV lane. *Transp. Res. Rec.* **1991**, 40–54.
17. Chirayil Nandakumar, S.; Mitchell, D.; Erden, M.S.; Flynn, D.; Lim, T. Anomaly detection methods in autonomous robotic missions. *Sensors* **2024**, *24*, 1330. [[CrossRef](#)]
18. Galvão, Y.M.; Castro, L.; Ferreira, J.; Neto, F.B.d.L.; Fagundes, R.A.d.A.; Fernandes, B.J. Anomaly detection in smart houses for healthcare: Recent advances, and future perspectives. *SN Comput. Sci.* **2024**, *5*, 136. [[CrossRef](#)]
19. Es-Swidi, A.; Ardchir, S.; Elghoumari, Y.; Daif, A.; Azouazi, M. Traffic congestion and road anomalies detection using CCTVs images processing, challenges and opportunities. In *International Conference on Advanced Intelligent Systems for Sustainable Development*; Springer: Berlin/Heidelberg, Germany, 2022; pp. 92–105.
20. Gao, J.; Zuo, F.; Ozbay, K.; Hammami, O.; Barlas, M.L. A new curb lane monitoring and illegal parking impact estimation approach based on queueing theory and computer vision for cameras with low resolution and low frame rate. *Transp. Res. Part A Policy Pract.* **2022**, *162*, 137–154. [[CrossRef](#)]
21. Kim, S.; Anagnostopoulos, G.; Barmponakis, E.; Geroliminis, N. Visual extensions and anomaly detection in the pNEUMA experiment with a swarm of drones. *Transp. Res. Part C Emerg. Technol.* **2023**, *147*, 103966. [[CrossRef](#)]
22. Yi, K.; Luo, K.; Chen, T.; Hu, R. An improved YOLOX model and domain transfer strategy for nighttime pedestrian and vehicle detection. *Appl. Sci.* **2022**, *12*, 12476. [[CrossRef](#)]
23. Wang, J.; Wang, X.; Hao, R.; Yin, H.; Huang, B.; Xu, X.; Liu, J. Incremental template neighborhood matching for 3D anomaly detection. *Neurocomputing* **2024**, *581*, 127483. [[CrossRef](#)]
24. Pan, Q.; Bao, Y.; Li, H. Transfer learning-based data anomaly detection for structural health monitoring. *Struct. Health Monit.* **2023**, *22*, 3077–3091. [[CrossRef](#)]
25. Yan, P.; Abdulkadir, A.; Luley, P.-P.; Rosenthal, M.; Schatte, G.A.; Grewe, B.F.; Stadelmann, T. A comprehensive survey of deep transfer learning for anomaly detection in industrial time series: Methods, applications, and directions. *IEEE Access* **2024**, *12*, 3768–3789. [[CrossRef](#)]
26. Bharambe, U.; Bhangale, U.; Narvekar, C. Role of multi-objective optimization in image segmentation and classification. In *Computational Intelligence in Image and Video Processing*; Chapman and Hall/CRC: Boca Raton, FL, USA, 2023; pp. 317–340.
27. Gonzalez, R.C. *Digital Image Processing*; Pearson Education India: Hoboken, NJ, USA, 2009.

28. Yuan, Y.; Huang, J.; Yu, J.; Tan, J.K.S.; Chng, K.Z.; Lee, J.; Kim, S. Application of machine learning algorithms for accurate determination of bilirubin level on in vitro engineered tissue phantom images. *Sci. Rep.* **2024**, *14*, 5952. [[CrossRef](#)] [[PubMed](#)]
29. Kilicaslan, M.; Tanyeri, U.; Demirci, R. Image retrieval using one-dimensional color histogram created with entropy. *Adv. Electr. Comput. Eng.* **2020**, *20*, 79–88. [[CrossRef](#)]
30. Mittal, H.; Pandey, A.C.; Saraswat, M.; Kumar, S.; Pal, R.; Modwel, G. A comprehensive survey of image segmentation: Clustering methods, performance parameters, and benchmark datasets. *Multimed. Tools Appl.* **2022**, *81*, 35001–35026. [[CrossRef](#)]
31. Vansh, V.; Chandrasekhar, K.; Anil, C.; Sahu, S.S. Improved face detection using YCbCr and Adaboost. In Proceedings of the 5th International Conference on Computational Intelligence in Data Mining (ICCIDM 2018), Burla, India, 15–16 December 2018; Springer: Berlin/Heidelberg, Germany, 2020; pp. 689–699.
32. Han, H.; Han, C.; Lan, T.; Huang, L.; Hu, C.; Xue, X. Automatic shadow detection for multispectral satellite remote sensing images in invariant color spaces. *Appl. Sci.* **2020**, *10*, 6467. [[CrossRef](#)]
33. Sahu, Y.; Tripathi, A.; Gupta, R.K.; Gautam, P.; Pateriya, R.K.; Gupta, A. A CNN-SVM based computer aided diagnosis of breast Cancer using histogram K-means segmentation technique. *Multimed. Tools Appl.* **2023**, *82*, 14055–14075. [[CrossRef](#)]
34. Kollem, S.; Reddy, K.R.; Rao, D.S. An optimized SVM based possibilistic fuzzy c-means clustering algorithm for tumor segmentation. *Multimed. Tools Appl.* **2021**, *80*, 409–437. [[CrossRef](#)]
35. Agrawal, S.; Natu, P. An improved Gaussian Mixture Method based background subtraction model for moving object detection in outdoor scene. In Proceedings of the 2021 Fourth International Conference on Electrical, Computer and Communication Technologies (ICECCT), Coimbatore, India, 15–17 September 2021; IEEE: Piscataway, NJ, USA, 2021; pp. 1–8.
36. Rakesh, S.; Hegde, N.P.; Gopalachari, M.V.; Jayaram, D.; Madhu, B.; Hameed, M.A.; Vankdothu, R.; Kumar, L.S. Moving object detection using modified GMM based background subtraction. *Meas. Sens.* **2023**, *30*, 100898. [[CrossRef](#)]
37. Zhou, Q.; Situ, Z.; Teng, S.; Chen, G. Comparative Effectiveness of Data Augmentation Using Traditional Approaches versus StyleGANs in Automated Sewer Defect Detection. *J. Water Resour. Plan. Manag.* **2023**, *149*, 04023045. [[CrossRef](#)]
38. Zuehlke, D.; Henderson, T.A.; McMullen, S. Machine learning using template matching applied to object tracking in video data. *Artif. Intell. Mach. Learn. Multi-Domain Oper. Appl.* **2019**, *11006*, 110061S.
39. Ge, Y.; Zhang, J.; Ren, X.; Zhao, C.; Yang, J.; Basu, A. Deep variation transformation network for foreground detection. *IEEE Trans. Circuits Syst. Video Technol.* **2020**, *31*, 3544–3558. [[CrossRef](#)]
40. Cao, Q.; Wang, Z.; Long, K. Traffic foreground detection at complex urban intersections using a novel background dictionary learning model. *J. Adv. Transp.* **2021**, *2021*, 3515512. [[CrossRef](#)]
41. Zhang, Y.; Zheng, W.; Leng, K.; Li, H. Background subtraction using an adaptive local median texture feature in illumination changes urban traffic scenes. *IEEE Access* **2020**, *8*, 130367–130378. [[CrossRef](#)]
42. Feng, J.; Zeng, D.; Jia, X.; Zhang, X.; Li, J.; Liang, Y.; Jiao, L. Cross-frame keypoint-based and spatial motion information-guided networks for moving vehicle detection and tracking in satellite videos. *ISPRS J. Photogramm. Remote Sens.* **2021**, *177*, 116–130. [[CrossRef](#)]
43. Yang, G.; Ramanan, D. Upgrading optical flow to 3d scene flow through optical expansion. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition 2020, Seattle, WA, USA, 13–19 June 2020; pp. 1334–1343.
44. Chen, X.; Jia, Y.; Tong, X.; Li, Z. Research on pedestrian detection and deepsort tracking in front of intelligent vehicle based on deep learning. *Sustainability* **2022**, *14*, 9281. [[CrossRef](#)]
45. Sun, M.; Davies, M.E.; Proudler, I.K.; Hopgood, J.R. Adaptive kernel Kalman filter. *IEEE Trans. Signal Process.* **2023**, *71*, 713–726. [[CrossRef](#)]
46. Chaurasiya, R.K.; Gondane, P.M.; Acharya, B.; Khan, M.I. Automatic road traffic analyzer using background subtraction, blob analysis, and tracking algorithms. In Proceedings of the 2023 7th International Conference on Computer Applications in Electrical Engineering-Recent Advances (CERA), Roorkee, India, 27–29 October 2023; IEEE: Piscataway, NJ, USA, 2023; pp. 1–6.
47. Sajeed, M.A.; Kelouwani, S.; Amamou, A.; Alam, M.Z.; Agbossou, K. Vehicle lane departure estimation on urban roads using GIS information. In Proceedings of the 2021 IEEE Vehicle Power and Propulsion Conference (VPPC), Gijon, Spain, 25–28 October 2021; IEEE: Piscataway, NJ, USA, 2021; pp. 1–7.
48. Momin, K.A.; Barua, S.; Jamil, M.S.; Hamim, O.F. Short duration traffic flow prediction using kalman filtering. *AIP Conf. Proc.* **2023**, *2713*, 040011.
49. Moghaddasi, S.S.; Faraji, N. A hybrid algorithm based on particle filter and genetic algorithm for target tracking. *Expert Syst. Appl.* **2020**, *147*, 113188. [[CrossRef](#)]
50. Chen, S.; Huang, L.; Chen, H.; Bai, J. Multi-lane detection and tracking using temporal-spatial model and particle filtering. *IEEE Trans. Intell. Transp. Syst.* **2021**, *23*, 2227–2245. [[CrossRef](#)]
51. Nissimagoudar, P.; Algur, N.; Bonageri, N.; Chavan, A.; Koppa, A.; Iyer, N.C. Multiple vehicle tracking using meanshift algorithm and 8-point connectivity. In Proceedings of the International Conference on Soft Computing and Pattern Recognition, Online, 14–16 December 2022; Springer: Berlin/Heidelberg, Germany, 2022; pp. 3–13.
52. Zhang, B.; Li, Z.; Perina, A.; Del Bue, A.; Murino, V.; Liu, J. Adaptive local movement modeling for robust object tracking. *IEEE Trans. Circuits Syst. Video Technol.* **2016**, *27*, 1515–1526. [[CrossRef](#)]
53. O'Mahony, N.; Campbell, S.; Carvalho, A.; Harapanahalli, S.; Hernandez, G.V.; Krpalkova, L.; Riordan, D.; Walsh, J. Deep learning vs. traditional computer vision. In Proceedings of the Advances in Computer Vision: 2019 Computer Vision Conference (CVC), Las Vegas, NV, USA, 2–3 May 2019; Springer: Berlin/Heidelberg, Germany, 2020; Volume 1, pp. 128–144.

54. Zhang, W.; Li, H.; Li, Y.; Liu, H.; Chen, Y.; Ding, X. Application of deep learning algorithms in geotechnical engineering: A short critical review. *Artif. Intell. Rev.* **2021**, *54*, 5633–5673. [[CrossRef](#)]
55. Long, G.; Zhang, Z. Deep encrypted traffic detection: An anomaly detection framework for encryption traffic based on parallel automatic feature extraction. *Comput. Intell. Neurosci.* **2023**, 3316642. [[CrossRef](#)] [[PubMed](#)]
56. Butt, U.M.; Ullah, H.A.; Letchmunan, S.; Tariq, I.; Hassan, F.H.; Koh, T.W. Leveraging transfer learning for spatio-temporal human activity recognition from video sequences. *Comput. Mater. Contin.* **2023**, *74*, 5017–5033.
57. Ouhami, M.; Hafiane, A.; Es-Saady, Y.; El Hajji, M.; Canals, R. Computer vision, IoT and data fusion for crop disease detection using machine learning: A survey and ongoing research. *Remote Sens.* **2021**, *13*, 2486. [[CrossRef](#)]
58. Carlson, M.P.; Bloom, I. The cyclic nature of problem solving: An emergent multidimensional problem-solving framework. *Educ. Stud. Math.* **2005**, *58*, 45–75. [[CrossRef](#)]
59. Vallenga, D.; Grypdonck, M.H.F.; Hoogwerf, L.J.R.; Tan, F.I.Y. Action research: What, why and how? *Acta Neurol. Belg.* **2009**, *109*, 81–90. [[PubMed](#)]
60. Farady, I.; Lin, C.-Y.; Chang, M.-C. PreAugNet: Improve data augmentation for industrial defect classification with small-scale training data. *J. Intell. Manuf.* **2024**, *35*, 1233–1246. [[CrossRef](#)]
61. Xu, M.; Yoon, S.; Fuentes, A.; Park, D.S. A comprehensive survey of image augmentation techniques for deep learning. *Pattern Recognit.* **2023**, *137*, 109347. [[CrossRef](#)]
62. Mathworks. Develop Apps Using App Designer. 2021. Available online: <https://www.mathworks.com/help/matlab/app-designer.html> (accessed on 10 April 2024).
63. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.
64. Sutskever, I.; Martens, J.; Dahl, G.; Hinton, G. On the importance of initialization and momentum in deep learning. In Proceedings of the International Conference on Machine Learning, Atlanta, GA, USA, 16–21 June 2013; PMLR: pp. 1139–1147.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.