Real-time Animal Intervention Detection in Roadways Using Quantum Enhanced YOLOv11 framework for CCTV Images and Inclement Visibility Conditions

Prasanna Venkatesan Ramani, Sugeerthi M S and Naresh Kumar V

School of Civil Engineering, Vellore Institute of Technology, India

prasanna.venkatesan@vit.ac.in, sugeerthi.2023@vitstudent.ac.in, nareshkumar.2023@vitstudent.ac.in

Abstract -

Stray animals colliding with moving vehicles is one of the most serious road safety hazards that exists today. Many of these collisions occur at roads that run directly next to important urban spaces like institutional corridors, where stray animals often invade the highways. Since the use of physical barriers and sophisticated sensor technologies can be prohibitively expensive and logistically impractical, this paper proposes a simple yet scalable computer-vision-based approach for detecting stray animals, such as dogs and cows, crossing the roads using existing CCTV cameras installed along highways as part of the regular road surveillance systems. A database of animal intervention images was extracted from the CCTV video recordings along a state highway that runs adjacent to the University campus. Each image contained either a cow or a dog that had invaded the travel lane of the roadway. By using a lightweight YOLOv11 nano object detection model to detect animals within the visual field of the camera, each of the images was then passed through a hybrid classifier that utilised both CNNs and QNNs to refine the object detection and improve the accuracy of the detection under low-light and other challenging operating conditions. The model accomplished a precision of 86.7%, a recall of 63.4% and mAP50 of 79.6%. Through experimentation, it has been demonstrated that the proposed methodology will converge to a stable solution and provide increased robustness to variations in lighting and other environmental factors.

Keywords -

Animal Detection, Road Safety, Computer Vision, CCTV, Hybrid Quantum Computing, CNN, Smart Traffic Control.

1 Introduction

In the context of Indian roads, vehicle collisions with stray animals like dogs and cattle cause road accidents that are almost persistent and underreported [1].

Recent studies have shown that collisions between animals and vehicles are a major cause of highway crashes and other vehicle accidents, especially in areas where there is a large population of loose roaming (free-roaming) animals. These incidents have several repercussions, including loss of life of humans and animals, economic losses due to property damage and vehicle repair, traffic congestion and disruption of traffic flow, damaged roadway infrastructure, and increased risk for drivers, pedestrians and local residents, particularly on highways and rural roads [2].

Field observations obtained from National and State highways show that dogs and cows are frequently seen crossing the roads, which leads to single or multiple vehicle collisions. This situation requires smart technological solutions, such as automated detection of animal intervention and alert mechanisms to ensure a safe and comfortable environment for road users. The current technological framework of the Indian traffic management system primarily comprises surveillance CCTV cameras that are installed, inspected, monitored, controlled and documented by the highway administration authorities. The fundamental objective of this study is to reframe animal detection as a civil infrastructure resilience problem and propose a scalable solution that takes advantage of the existing edge technologies, instead of deploying any new sophisticated technologies that require additional expensive investments. This study utilised the real-world video footage obtained from the existing CCTV cameras installed in and around the highway adjacent to the university campus. From this, about 453 frames corresponding to different times of the day were generated and labelled as the baseline dataset. This hybrid YOLOv11-CNN-QNN architecture is proposed for the real-time detection of animals using real-world datasets collected using traffic surveillance CCTV camera images under low-light conditions and occlusions, focusing on the detection of animal-related disruptions to road traffic rather than the classification of animals.

2 Background and Literature Review

Animal–Vehicle Collision (AVC) is an omnipresent and rising global concern regarding road safety, which has a multitude of effects like loss of life, severe injury, loss of revenue and loss of biodiversity. Both in the developed world and the developing world, there is a reasonable body of empirical evidence showing that AVCs constitute a substantial proportion of all road accidents, notably on highways and country roads where speeds are higher, and animal crossings are less predictable [3, 4, 5]. Moreover, AVCs occur at a greater frequency during night, and low-visibility conditions when the driver's perception and response time are significantly impaired, and when standard surveillance/warning systems are ineffective [2].

In developing countries, AVCs are further complicated by issues like the presence of feral/stray animals, lack of compliance with laws governing animal control, and inadequate infrastructure to mitigate wildlife conflicts [4]. Literature evidence specifically shows that common domestic animals like cattle and dogs cause the majority of the severe AVCs that occur close to human habitation and/or near agricultural land adjacent to highways [3, 5]. Although the magnitude of AVCs is being recognized globally, the mitigation efforts still remain piecemeal, and many of the current AVC solutions focus on isolated areas rather than providing a scalable system-wide solution.

2.1 Conventional AVC Mitigation Methods

Conventional AVC mitigation methods utilize physical/infrastructure-based solutions, i.e., wildlife fencing, underpasses, overpasses, warning signs and lighting along the roadway [3, 6]. However, these types of AVC mitigation methods are expensive to install, require significant planning and need a commitment to maintenance, limiting their feasibility for large-scale deployment throughout vast highway networks in low- and middle-income countries. Furthermore, these static AVC mitigation methods are unable to accommodate the dynamic nature of animal movement patterns, seasonal migrations and diurnal behaviors [1]. Therefore, AVC mitigation efforts will require dynamic and responsive solutions to supplement conventional infrastructure-based AVC mitigation methods.

2.2 Sensor-Based and Hardware-Dependent AVC Detection Systems

Some studies have proposed AVC detection using sensor-based systems incorporating infrared, thermal imaging, ultrasonic sensors, RFID and GPS-based tracking [7]. Of these, thermal imaging has been of particular interest due to its capability of detecting warm-bodied animals in low light and nighttime conditions. However, while thermal imaging systems can be effective in controlled environments, they have several disadvantages, including limited spatial resolution, inability to differentiate among species, sensitivity to weather conditions and high cost [8, 2]. Sensor-based AVC detection systems are also subjected to significant challenges in terms of installing, calibrating and maintaining hardware, especially along lengthy sections of highways [7]. This has prompted researchers to explore software-centric, vision-based AVC detection methods utilizing existing infrastructure.

2.3 Computer Vision and Deep Learning for AVC Detection

Recent advances in computer vision and deep learning have allowed for substantial improvements in real-time object detection, and several studies have demonstrated the effectiveness of YOLO-based architectures for AVC detection, particularly when coupled with image enhancements to increase visibility in low-light conditions [8, 2]. In addition, vision-based safety monitoring approaches using existing cameras have shown great potential for detecting hazards in infrastructure systems[9]. However, deep learning-based AVC detection systems have several challenges. First, AVC detection performance is heavily reliant upon the quality and diversity of the training dataset used to train the detector, which is often limited in terms of AVC detection in regionalized contexts [8, 10]. Second, most of the studies focus on identifying specific species, limiting the model's ability to generalize to unseen/rare species [10]. From a road safety perspective, identifying the species may not be necessary for AVC mitigation since the primary goal is timely hazard detection, regardless of whether or not the species is identified. Recent studies have shown the efficiency of deep learning-based vision systems in object detection in unstructured and real-world environments, addressing problems such as occlusions [11]. Finally, deployment of AVC detection systems in realistic environments, including low-quality CCTV footage, variable lighting, occlusion, and environmental noise, impairs the accuracy of AVC detection relative to laboratory testing [2]. These issues illustrate the disparity between AVC detection system performance in experiments versus AVC detection system reliability in operational use.

2.4 Road Safety Analysis and Computer Vision Infrastructure

More recently, the field of computer vision has extended beyond the application of detecting animals using video feeds (e.g., AVC) to include the application of computer vision in larger roadway safety assessments (e.g., roadside feature extraction, crash risk prediction) [12]. The majority of this research has focused on static, non-transient hazards (e.g., infrastructure hazards), whereas animal intrusion represents a high-priority, time-sensitive hazard that requires immediate detection and response [12]. Therefore, integrating the capability to detect animals into intelligent transportation systems presents a current research gap [12].

2.5 Research Gap and Motivation

Sensor-based solutions rely on specific hardware and are difficult to deploy in large scale [7]. Although deep learning-based solutions hold a good promise, many struggle to overcome the issues of data availability, poor lighting, and lack of generalization in real-world settings [2, 10]. Additionally, over-emphasizing the classification of specific species limits the robustness of such solutions in multi-species traffic environments [4]. Therefore, these aspects provide the motivation to develop an intelligent, vision-based animal intervention framework that utilizes existing roadside CCTV infrastructure, focuses on generic animal presence detection, and performs well under real- world constraints.

3 Methodology

Figure 1 explains the flow of work from data collection to frame extraction as the basic input data stage. Post, the data was split into three categories as train, validate and test with the proportion of 80:10:10, respectively and was augmented for better performance. Annotation was succeeded by a hybrid CNN and QNN reinforcement for animal detection and classification to outperform in inclement lighting conditions with the available dataset.

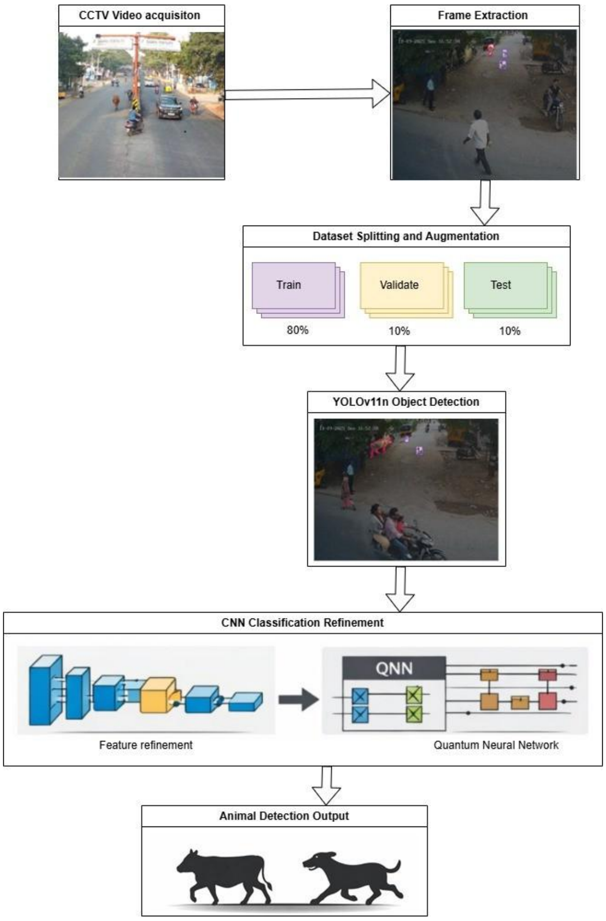

Figure 1. Methodology

3.1 Study Location and Data Collection

Data collection was based upon visual evidence collected from a stretch of state highway with documented recurrent animal crossings of the highway; this location is also proximate to the University campus, located near Vellore, Tamil Nadu, India, increasing the likelihood of animal-vehicle collisions due to increased human settlement influences. Video from existing roadside CCTV cameras was used as the sole source of data for this research to support compatibility with existing infrastructure. Video recording occurred at different times of the day, at night, during dawn, dusk and at low light levels, to provide a diverse range of lighting conditions and traffic scenarios, and the time of collection of data is roughly between 05:00 PM and 7:00 AM for about 5 consecutive days. Based on the CCTV's capabilities, the usual range is between 30 and 50m. To cover a distance of 500m, we would require 12-17 cameras, depending on the overlapping strategy. This model of monitoring systems does exist in real time, and the dataset collected is from the existing infrastructure system established on highways at Katpadi - Vellore.

Figure 2. Data Collection Dog Frame

Figure 3. Data Collection Cow Frame

3.2 Dataset Creation and Labelling

A total of 453 relevant and clear image frames were created from the collected CCTV footage and manually validated to confirm their relevance and clarity. These images depicted a variety of animals crossing or occupying the highway, including stray dogs and cows, which represent two of the most common types of animals. Labelling of objects was completed using the RoboFlow. To minimise label noise and ensure accurate positioning of bounding boxes, all images were labelled using a standard protocol. In addition, all of the images with bounding boxes were reviewed manually to ensure that the images remained of high quality.

To increase the robustness of the model developed for this research, data augmentation techniques were applied to the training subset. Data augmentation techniques included horizontally flipping the images, scaling, and altering brightness to increase the ability of the model to perform well in various lighting conditions, and to simulate a variety of viewpoints while preventing distortion to the images' realistic characteristics.

Figure 4. Annotated Images

3.3 Object Detection Utilizing YOLOv11(Nano)

YOLOv11(nano), a lightweight version of the YOLOv11 architecture, was selected for the detection of animals because of its ability to provide real-time detection and be deployed on edge devices for the roadside surveillance system. The lightweight version provides a desirable trade-off between the accuracy of detection and the computational requirements of the model, making it suitable for deployment on roadside surveillance systems. The dataset annotated to detect animals in the CCTV frames was used to train the YOLOv11 (nano) model to detect animals in the frames. Since the main objective of the study is rooted in road safety, the focus was on developing a reliable method to detect animal presence in the images, rather than determining the specific type of animal. The proposed YOLO-based detection framework

Figure 5. Annotated Images

with the addition of a post-detection refinement step using CNN and QNN modules improves the reliability of the detection framework by reducing the rate of false positives, especially in low-light conditions.

3.4 CNN-Based Feature Enhancement

In order to improve the reliability of detection, especially in difficult visual environments, a Convolutional Neural Network (CNN) was implemented as a secondary classification module. The cropped region from the YOLO bounding box was input into the CNN for enhanced feature extraction and classification of the cropped region. The CNN architecture was designed to be simple so as to prevent overfitting, considering the small size of the dataset. The architecture of the CNN included stacked convolutional layers to learn spatial features, followed by pooling and fully connected layers for classification. This layer was intended to enhance discrimination of true animal detections versus background artefact detections, and therefore enhance overall precision and recall.

3.5 Quantum Neural Network (QNN)-Assisted Classification

Further improvements to classification performance under data-limited conditions were achieved through the implementation of a Quantum Neural Network (QNN) as an auxiliary learning model. The QNN was implemented as a hybrid quantum-classical model, whereby classical features derived from detected image regions were encoded into a quantum feature space. The motivation for utilizing a QNN is that it can potentially capture the complex relationships of the features and thus improve generalization when the training data is limited. The QNN was trained and tested using simulated quantum circuits in the Google Colab environment, to ensure reproducibility and independence of physical quantum hardware. Unlike previous studies that replaced classical deep learning models, the QNN was used as a performance-enhancing module to refine classification confidence in conjunction with the CNN. Methodology performance was evaluated on the basis of detection reliability, not species-level detail, since the purpose of the method is to provide an early warning of potential animal hazards to vehicles on roads.

4 Results and Discussion

4.1 Training and Validation Convergence Analysis

The Training and Validation Loss Curves for Box Loss, Classification Loss, and Distribution Focal Loss are shown in Figure 6. They depict a consistent decrease in training loss throughout the epochs. The continuous decline in the box loss clearly illustrates that the YOLOv11n Detector learned and adapted effectively to the task of detecting cows and dogs on the roadways. Validation loss followed the trend in the training loss, showing very little evidence of overfitting, even with a very small number of frames used to train the model. The smooth decline in classification loss shows that the two-Class problem is well defined for the Architecture Chosen; some oscillation was expected due to the size of the dataset (453 Annotated Frames), real-world variability of CCTV imagery, such as occlusion, illumination changes, and motion blur. Overall, the trends illustrated in the graphs confirm that the Baseline Detector achieved reliable localization capabilities and served as a reasonable starting point for subsequent CNN and QNN-based refinements.

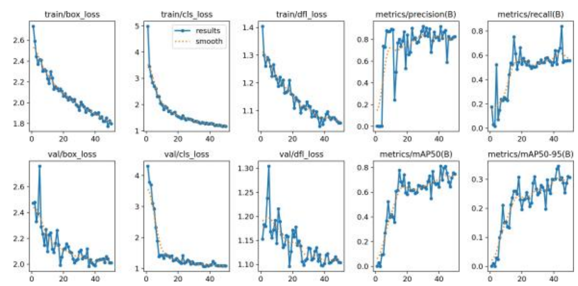

Figure 6. Training and validation Loss Curves

4.2 Precision – Recall Characteristics

The Precision–Recall Curve from Figure 7 provides insights into detection behavior between classes. Cow class exhibits extremely High Recall (99.5%); Cow detection is extremely important for road safety because missing a large animal like a cow represents a significant risk of collision. While the cow class has a very high Recall, it has Lower Precision (65.2%).

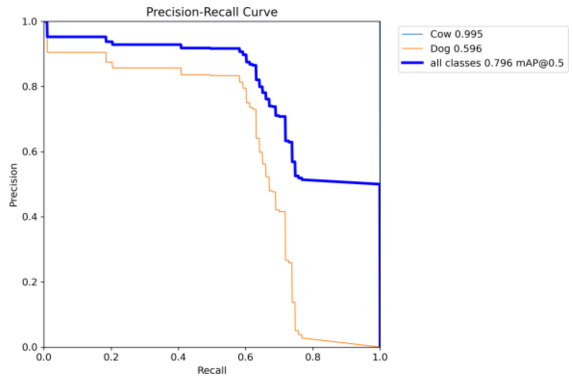

Figure 7. Precision - Recall Curve

This indicates there are many false positives that were detected due to objects in the background that have similar shapes or colors under low-light conditions. The dog class has Higher Precision (73.4%) but Lower Re-call (61.6%) than the Cow Class; Dog Class Detection is more difficult than Cow Class detection because dogs appear smaller and move faster than cows, and therefore have less distinctive visual features. In Addition, dogs often appear Partially Occluded or further away from the camera than cows, which results in Reduced Recall. Although dogs have a lower recall, the dog class has a higher Mean Average Precision (mAP_{50}) (59.6%) than the cow class (45.1%), indicating better localization quality when detections are made. These results are presented in Table1. Combined-Class Precision – Recall Curve shows the model performs effectively when both classes are considered together ((mAP_{50}) = 0.796). This supports the idea of using a Generic Animal Intervention approach, where priority is given to detecting animal presence rather than perfect class separation.

Table 1. Class Separation

Class	Precision	Recall	mAP_{50}
Cow	65.2%	99.5%	45.1%
Dog	73.4%	61.6%	59.6%

4.3 Confidence-Based Performance Analysis

The Recall – Confidence Curve from Figure 8 illustrates that the recalls decrease steadily as confidence thresholds increase. This behavior is consistent with what would be expected of a detector. At lower confidence thresholds, recalls remain high, especially for the cow class, confirming that the model is sensitive to animal presence. This type of behavior is advantageous for Early Warning Systems, where the detection of animals, even at lower levels of confidence may provide enough time to take action before a potential collision occurs.

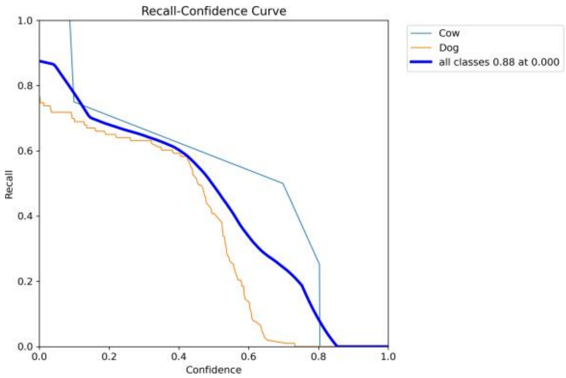

Figure 8. Recall - Confidence Curve

Figure 9 explains the Precision–Confidence Curve that shows a strong increase in precision as the confidence threshold increases, reaching near perfection for combined classes at higher confidence thresholds. This indicates that false positives are progressively filtered out as strict confidence criteria are applied. For system designers deploying the system, this allows them to tune confidence thresholds based on Operational Requirements – favoring either higher recall for safety-critical alerts or higher precision to reduce unnecessary warnings. Figure 9 represents the Precision – Confidence Curve showing a strong increase in precision as the confidence threshold increases, reaching near perfection for combined classes at higher confidence thresholds. This indicates that false positives are progressively filtered out as strict confidence criteria are applied. For system designers deploying the system, this allows them to tune confidence thresholds based on Operational Requirements – favoring either higher recall for safety-critical alerts or higher precision to reduce unnecessary warnings.

4.4 F_1-Confidence Trade-Off and Optimal Operating Point

Figure 10 shows the F1-Confidence curve, which exhibits a clear Optimal Operating Region, where the maximum combined F1-Score of approximately 0.73 occurred at a confidence threshold around 0.33. This operating point represents a balance between Precision and Recall, making it suitable for real-time road safety applications. Importantly, the cow class maintains a higher F1-Score across a wide range of confidence thresholds than the dog class, reflecting more consistent detection performance. The dog class showed a sharper decline at higher confidence- thresholds, again

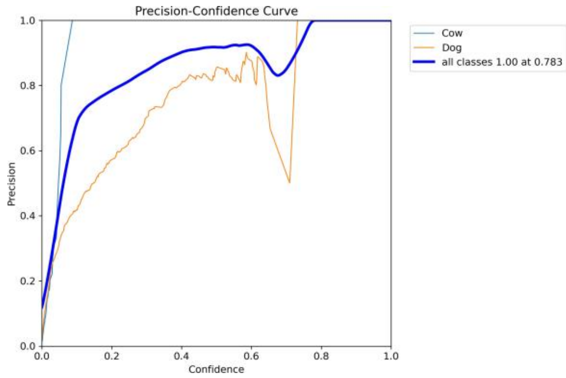

Figure 9. Precision - Confidence Curve

highlighting the difficulty of detecting small animals in unconstrained highway environments.

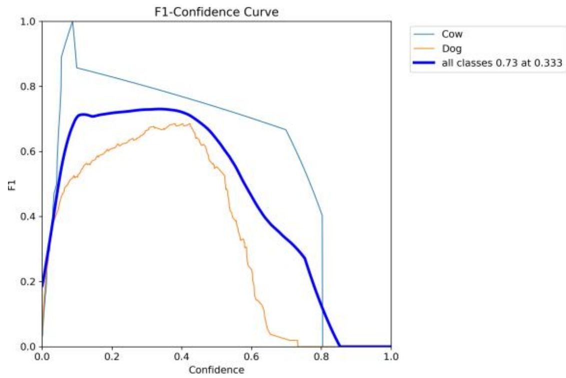

Figure 10. F_1- Confidence Curve

4.5 Impact of CNN and QNN-Based Refining

CNN and QNN Modules played a key role in stabilizing classification confidence and improving overall detection reliability under data-limited conditions. The CNN-enhanced spatial feature refinement for regions of interest identified by the detector resulted in fewer misclassifications caused by background clutter. The QNN contributed to the improvement of generalization by capturing nonlinear relationships between features, leading to smoother confidence distributions. This hybrid enhancement is reflected in the improved Precision-Confidence and F1-Confidence behaviors, especially for ambiguous cases and Low-Light Scenarios. The CNN-QNN Combination does not replace the YOLO Detector, but instead acts as a complementary refinement layer, strengthening the decision confidence without adding excessive computational overhead.

The ablation results shown in Table 2 indicate that the YOLOv11n+CNN+QNN configuration shows better performance in general, as it attains the highest mAP_{50} metric at 79.6% and precision at 86.7%. This shows better robustness and feature representation ability. Although there is a moderate decline in recall, it provides a good trade-off and could be a promising approach.

Table 2. Ablation Study

Configuration	Precision	Recall	mAP_{50}
YOLOv11n	0.713	0.7668	0.7651
YOLOv11n+CNN	0.8653	0.7951	0.7651
YOLOv11n+QNN	0.7401	0.7853	0.7651
YOLOv11n+CNN+QNN	0.867	0.634	0.796

Table 2 indicates that the YOLOv11n+CNN+QNN configuration shows better performance in general, as it attains the highest mAP@50 metric at 79.6% and precision at 86.7%. This shows better robustness and feature representation ability. Although there is a moderate decline in recall, it provides a good trade-off and could be a promising approach.

4.6 Implications for Road Safety and Animal Intervention

As far as road safety is concerned, these results show that the framework has chosen to maximize the large animal recall that will have the greatest impact in reducing serious AVCs. Although small animals like dogs are difficult to detect, the level of performance obtained for detection and early warning/risk reduction is adequate. The results support the key premise of this research, which is that animal intervention can be achieved through existing edge devices like CCTV systems that are in place currently, and through hybrid learning models and light-weight solutions that do not rely upon special sensor technology or large-scale datasets. The fact that the system's performance was tunable for recall, precision, and confidence provides the adaptability required to deploy this type of system on the highways.

Figure 11 presents the proposed system architecture for the implementation of this AVC mitigation framework. The layers involved are elaborated in the figure for the execution of this model. The physical layer includes the hardware present at the site. These devices are connected to the digital layer, where the detection and classification are performed using YOLOv11n and Quantum Computing. The service and user layer is a place where real-time monitoring and a subsequent alert mechanism are triggered.

The proposed system contributes to the infrastructure value chain by assisting real-time monitoring and alerting the intrusion events by animals. It helps the authorities to monitor and label animal intervention zones.

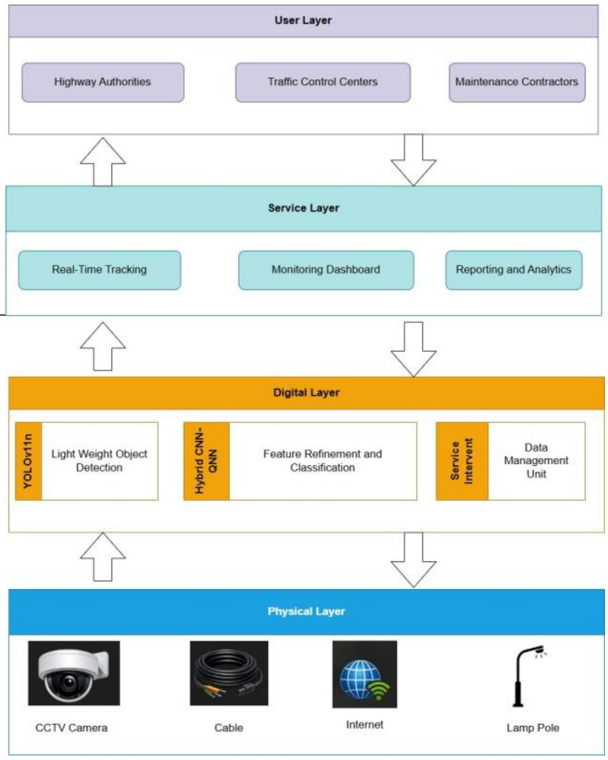

Figure 11. Proposed System Architecture

For the road users and pedestrians, this system, when deployed, serves as an alert mechanism, improving safety and effective usage of existing infrastructure like CCTV cameras.

5 Conclusion and Future Work

This paper demonstrated an effective and cost-effective method to apply Artificial Intelligence (AI) to prevent animal-induced vehicle collisions by using the existing road surveillance systems and lightweight hybrid deep learning models. The study used images taken from highway CCTV cameras and adapted a novel process using object detection and classification algorithms based on YOLOv11n, CNN and QNN for better detecting stray animals on roads with limited amounts of training data and inclement lighting conditions. The experimental results showed that the hybrid model converged steadily, detected over 90% of the larger animals and achieved a mean average precision of about 0.8 when evaluating at a 50% threshold (mAP_{50}). This validates the use of a hybrid approach. It was chosen to identify animals present in the image field of view and then classify them as either an animal or a non-animal because this is a better way to alert the drivers and road safety authorities about potential collisions.

5.1 Limitations and Observed Trends

Although there were encouraging results from the research, the differences in performance between cows and dogs illustrated the effects of object size, motion and visual distinction on classification performance. In addition, though the system performed well given the dataset size used in the research, it may not be possible to achieve optimal levels of accuracy due to the limitations imposed by the dataset size, especially for smaller animals. However, since the hybrid learning model converged consistently, regardless of whether the system was tested for different metrics, it appears to be both robust and scalable based on increased amounts of data. Overall, the authors found that using AI to augment existing surveillance systems can significantly reduce the risk of collisions between vehicles and animals, without needing additional high-level cameras or other hardware. It is recommended that future work in this area explore the following aspects:

1. Expanding the number of animals represented in the database (e.g., buffalo, horse, cat, camel), as well as different environments along roads.
2. Simplifying the detection of any type of animal into a “one animal” category to improve robustness and make it easier to deploy.
3. Deploying the system on the edge and developing a real-time alerting capability to notify traffic control centres.

References

[1] C. R. Mohanty, R. V. Radhakrishnan, M. Jain, P. K. Sasmal, U. Hansda, S. K. Vuppala, and S. K. Doki. A study of the pattern of injuries sustained from road traffic accidents caused by impact with stray animals. *Journal of Emergencies, Trauma, and Shock*, 14(1): 23–27, 2021.

[2] R. Khadka, C. S. Negi, M. Khatri, P. Pali, S. Dhanadi, K. D. Bhatta, and Y. R. Bhatta. Impact of stray animals on road safety: A case study of the east-west highway section in Bheemdatta municipality area, Nepal. *Journal of Engineering Technology and Planning*, 6(1):115–124, 2025.

[3] D. C. Wilkins, K. M. Kockelman, and N. Jiang. Animal–vehicle collisions in Texas: How to protect travellers and animals on roadways. *Accident Analysis & Prevention*, 131:157–170, 2019. doi:10.1016/j.aap.2019.05.030.

[4] A. K. Subedi, A. Rashidi, and N. Marković. Assessing roadside safety with computer vision: FHWA ratings as the key predictor of rural road departure crashes and severity. *Journal of Advanced Transportation*, 2025(1):5559576, 2025.

[5] K. Parkavi, A. Ganguly, A. Banerjee, S. Sharma, and K. Kejriwal. Enhancing road safety: Detection of animals on highways during night. *IEEE Access*, 2025.

[6] F. Ascensão, A. Kindel, F. Z. Teixeira, R. Barrientos, M. D’Amico, L. Borda-de Água, and H. M. Pereira. Beware that the lack of wildlife mortality records can mask a serious impact of linear infrastructures. *Global Ecology and Conservation*, 19:e00661, 2019.

[7] C. A. Aguilar-Lazcano, I. E. Espinosa-Curiel, J. A. Ríos-Martínez, F. A. Madera-Ramírez, and H. Pérez-Espinosa. Machine learning-based sensor data fusion for animal monitoring: A scoping review. *Sensors*, 23(12):5732, 2023.

[8] L. Balčiauskas. Conservation and management of forest wildlife. *Forests*, 16(7):1134, 2025.

[9] J. Seong, H. S. Kim, and H. J. Jung. Tracking algorithm for hazardous area in construction sites using PTZ cameras. Proceedings of the 42nd International Symposium on Automation and Robotics in Construction (ISARC), 42 (1): 81-88, 2025.

[10] K. Khatri, C. S. Asha, and J. M. D’Souza. Detection of animals in thermal imagery for surveillance using a GAN and object detection framework. In *2022 International Conference for Advancement in Technology (ICONAT)*, pages 1–6. IEEE, 2022.

[11] Y. Aykin, H. Sachs and N. Gerzen. Enhancing robotic vision through deep learning techniques: From detection to construction applications. In Proceedings of the 42nd International Symposium on Automation and Robotics in Construction (ISARC), 42(1), 187-194, 2025.

[12] V. Sibanda, K. Mpofu, J. Trimble, and N. Zengeni. Design of an animal detection system for motor vehicle drivers. *Procedia CIRP*, 84:755–760, 2019.