Published by :
https://www.ijert.org/
An International Peer-Reviewed Journal

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 02, NCRTCS - 2026

A Smart Urban Traffic Framework for Emergency Vehicle Priority Using Audio-Visual Deep Learning

Santosh D. Pandure
Department of Computer Science and Information Technology,
Dr. Babasaheb Ambedkar Marathwada University,
Chhatrapati Sambhaji Nagar, India.

Dr. Pravin L. Yannawar
Department of Computer Science and Information Technology,
Dr. Babasaheb Ambedkar Marathwada University,
Chhatrapati Sambhaji Nagar, India.

Abstract **-** This research work proposes an advanced and intelligent traffic management framework designed specifically to support the smooth movement of emergency vehicles in congested urban environments. The central objective of the system is to reduce delays faced by ambulances, fire trucks, and police vehicles by granting them priority at traffic intersections. To achieve this, the approach employs deep learning–based models capable of accurately detecting emergency vehicles in real time and dynamically controlling traffic signal operations along their travel route. The system identifies emergency vehicles through the combined use of audio and visual data, which significantly improves detection reliability in noisy and visually complex city conditions. Audio signals such as sirens are analyzed alongside video feeds from traffic cameras to confirm the presence of emergency vehicles. By integrating multiple artificial intelligence–driven algorithms, the proposed method minimizes false detections and ensures consistent performance across varying traffic densities and environmental conditions. Once an emergency vehicle is detected, the system intelligently modifies the sequence and timing of traffic signals at nearby intersections to create a clear path, allowing the vehicle to pass with minimal interruption. This adaptive signal control mechanism not only shortens emergency response times but also helps maintain smoother traffic flow for other road users by preventing unnecessary congestion. Overall, the proposed solution contributes to the development of smarter urban transportation systems by enhancing emergency response efficiency, improving road safety, and optimizing traffic signal coordination. The findings of this research demonstrate the potential of AI-enabled traffic control systems to address critical challenges in modern urban mobility.

Keywords— Emergency Vehicle Detection, Deep Learning, Multi-Modal Learning, Traffic Signal Control, Smart Cities.

I. INTRODUCTION

Effective traffic control in cities plays a vital role in ensuring that emergency services can respond without delay. When emergency vehicles are slowed down by traffic congestion or poorly coordinated signals, the impact can be severe and life-threatening. This study focuses on improving traffic signal control strategies to give priority to emergency vehicles in urban settings. The main aim of the research is to design an intelligent system that uses deep learning methods to detect emergency vehicles and grant them signal priority at intersections along their travel path. By dynamically adjusting traffic signals, the proposed system seeks to enhance the speed and reliability of emergency responses in city environments.

II. LITERATURE REVIEW

Recent research has explored the use of deep convolutional neural networks for detecting emergency vehicles in heavy traffic scenarios, demonstrating improved recognition accuracy and reduced response delays. Vision-based surveillance systems using ConvNet architectures have proven effective in identifying emergency vehicles under congested conditions [1]. With the increasing deployment of traffic cameras in cities, camera-based vehicle detection has become a dominant approach. Studies emphasize that accurate vehicle segmentation requires effective handling of shadows, which are classified into multiple categories and must be removed to avoid detection errors [2]. Sound-based analysis has gained attention in smart city research due to its ability to capture emergency events beyond visual range. However, urban sound environments are highly unstructured, making classification challenging. Deep learning models, combined with feature extraction and data augmentation, have been shown to significantly enhance environmental sound recognition accuracy [3]. Several works introduce intelligent sound event detection frameworks for autonomous and smart vehicles. By applying signal enhancement techniques such as gammatone filtering and STFT, these systems achieve better classification of emergency sounds and demonstrate potential for future multi-modal expansion [4]. Object detection research has extensively compared CNN-based models such as SSD, MobileNet, and Faster R-CNN. Findings indicate that lighter models offer faster inference, while deeper architectures provide higher accuracy, making model selection application-dependent [5]. Comprehensive reviews categorize modern object detection techniques into anchor-based, anchor-free, and transformer-based models. These studies highlight that targeted data augmentation plays a crucial role in improving detection performance [6,7]. Audio classification studies, including bird sound recognition and environmental sound detection, demonstrate that deep learning fusion techniques can deliver strong performance with high accuracy and balanced F1 scores [8]. Fast R-CNN is widely recognized for achieving superior detection efficiency compared to earlier object detection frameworks [9]. Automatic audio event detection has been successfully applied in safety-critical applications. Deep learning models using MFCC and log-Mel

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 02, NCRTCS - 2026

features show promising scalability and accuracy, especially when trained on large datasets [10]. In emergency vehicle detection, CNNs and transfer learning approaches have been used to identify vehicles in real-time traffic conditions, enabling the creation of green corridors and reducing accident risks [11]. Alarm and siren detection research demonstrates that supervised learning models can achieve high detection accuracy with minimal false alarms, even in noisy urban environments, when appropriate preprocessing techniques are applied [12,13]. Intelligent traffic control systems driven by deep learning dynamically adjust signal timings based on traffic density and emergency vehicle presence, offering an adaptive solution to congestion management [14]. Image-based ambulance detection systems using neural networks have shown potential in clearing traffic paths for emergency vehicles, thereby improving survival rates during emergencies [15]. Recent studies propose hybrid emergency vehicle detection systems that integrate visual and acoustic cues. These systems achieve higher robustness and accuracy in real-world conditions compared to single-modality approaches [16]. Advancements in CNN training strategies have improved image classification performance, with deep models consistently outperforming traditional classifiers [17]. Automated emergency vehicle detection using CCTV footage and deep learning has been demonstrated to be effective in congested urban environments, significantly improving detection reliability [18,19]. Acoustic-based emergency vehicle detection using ensemble deep learning models achieves exceptionally high accuracy and supports intelligent traffic management systems [20]. Environmental sound classification using CNN architectures such as AlexNet and GoogLeNet confirms the suitability of image-based deep learning models for audio spectrogram analysis [21]. Hybrid sensor systems combining magnetic and acoustic data have also been explored for traffic monitoring, with promising results for emergency vehicle identification [22].

III. SYSTEM ARCHITECTURE AND PROPOSED METHODOLOGY

The proposed system is designed as a multi-stage framework that integrates emergency vehicle detection, decision fusion, and traffic signal control. Figure representations and workflow diagrams may be incorporated during final publication.

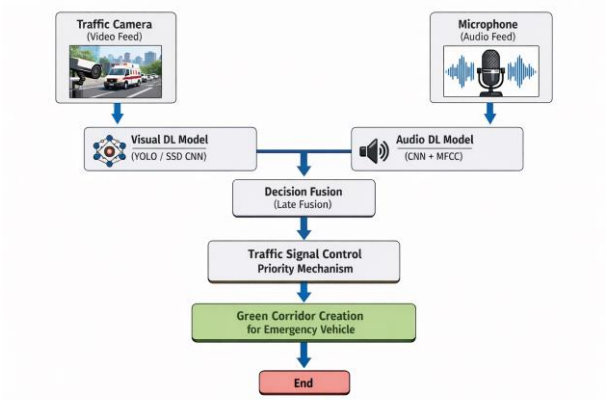

Figure 1. Multi-modal emergency vehicle detection and traffic signal prioritization framework.

A. Overall System Design

The system operates by continuously monitoring traffic conditions using roadside cameras and microphones installed near intersections. Visual and acoustic data streams are processed independently using dedicated deep learning models. The outputs of these models are then combined using a decision fusion mechanism to determine the presence of an emergency vehicle. Once detected, the traffic signal controller is triggered to prioritize the corresponding lane.

B. Visual-Based Emergency Vehicle Detection

Visual detection focuses on identifying emergency vehicles from real-time video feeds. The process begins with data collection from traffic cameras under different lighting, weather, and traffic conditions. Emergency vehicle instances are manually annotated to create a labeled dataset. To improve model generalization, data augmentation techniques such as rotation, scaling, and brightness adjustment are applied. Deep learning-based object detection models, including YOLO and SSD, are employed due to their balance between speed and accuracy. The models are trained and fine-tuned using the annotated dataset, and their performance is evaluated using standard metrics such as precision, recall, and F1-score. Post-processing techniques, including confidence thresholding and non-maximum suppression, are applied to refine detection results. The trained model is then deployed for real-time inference at traffic intersections.

C. Audio-Based Emergency Vehicle Detection

Audio-based detection complements visual analysis by identifying emergency sirens that may not be visible due to occlusion or poor lighting. Audio data consisting of siren sounds from ambulances and fire trucks, along with ambient traffic noise, is collected and labeled accordingly. Preprocessing steps include noise filtering, normalization, and feature extraction. Acoustic features such as Mel-Frequency Cepstral Coefficients (MFCCs) and spectrogram representations are extracted to capture distinctive siren characteristics. Deep learning models such as Convolutional Neural Networks (CNNs) are trained on these features to classify audio segments. The trained audio model is capable of performing real-time inference, allowing the system to detect approaching emergency vehicles even before they appear in the camera's field of view.

D. Decision Fusion Strategy

To enhance robustness, a decision-level fusion approach is adopted. Instead of merging raw features, the final predictions from the visual and audio models are combined. If either model identifies the presence of an emergency vehicle, the system confirms detection. This late fusion strategy improves system reliability, especially in challenging scenarios where one modality may fail due to environmental factors. The fusion mechanism can be further refined based on real-world deployment feedback.

E. Traffic Signal Prioritization Mechanism

Once an emergency vehicle is detected, the traffic signal controller dynamically adjusts the signal phases. The green signal duration for the corresponding lane is extended, while conflicting directions are temporarily halted. This creates a

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 02, NCRTCS - 2026

clear and uninterrupted path for the emergency vehicle. The prioritization logic is designed to minimize disruption to overall traffic flow while ensuring emergency vehicles receive immediate right-of-way. After the emergency vehicle passes, the system gradually restores normal signal operation.

IV. EXPERIMENTAL SETUP AND DATA DESCRIPTION

A. Dataset Description

Two distinct datasets were used in this study. The visual dataset consists of 2352 labeled images collected from Kaggle, representing emergency and non-emergency vehicles. The audio dataset includes 3-second WAV recordings of ambulance sirens, fire truck sirens, and general traffic noise. Each category contains 200 samples.

B. Data Preprocessing

Image preprocessing involves resizing, normalization, noise reduction, and artifact removal. Libraries such as NumPy and PIL are used to manage pixel-level operations efficiently. For noise reduction, Non-Local Means Denoising techniques are applied to enhance image clarity. Audio preprocessing includes silence removal, normalization, and feature extraction. Spectrogram and MFCC representations are generated to serve as input for the audio classification model.

V. MODEL TRAINING AND IMPLEMENTATION

The CNN architecture used in this study consists of convolutional layers for feature extraction, pooling layers for dimensionality reduction, and fully connected layers for classification. Batch processing is employed to efficiently handle large datasets and optimize memory usage. The models are trained using supervised learning techniques, and back-propagation is applied to minimize classification error. Hyper-parameters such as learning rate, batch size, and number of epochs are tuned experimentally.

VI. PERFORMANCE EVALUATION AND RESULTS

The image-based detection model achieves high classification performance, with an overall accuracy of 96%. Precision and recall values remain consistently high across emergency vehicle categories. The audio-based model demonstrates superior performance, achieving an overall accuracy of 99%. High F1-scores indicate balanced precision and recall, confirming reliable siren detection even in noisy environments. The fusion-based system further improves robustness, ensuring emergency vehicle detection under diverse urban conditions.

VII. DISCUSSION

The experimental results highlight the effectiveness of combining visual and audio modalities for emergency vehicle detection. While visual detection performs well under clear conditions, audio-based detection proves valuable in low-visibility scenarios. The fusion strategy ensures system reliability and reduces false negatives. The proposed traffic signal prioritization mechanism successfully minimizes emergency vehicle delays without significantly impacting normal traffic flow. This makes the system suitable for real-world urban deployment.

VIII. CONCLUSION

This research presents an intelligent traffic signal prioritization system that leverages deep learning and multi-modal data to improve emergency vehicle mobility in urban environments. By integrating real-time detection with adaptive signal control, the system reduces response delays and enhances road safety. The results demonstrate strong detection accuracy and practical feasibility for smart city applications.

IX. FUTURE WORK

Future research will focus on large-scale real-world deployment, integration with GPS-based routing, and reinforcement learning-based signal optimization. Expanding the system to handle multiple simultaneous emergency vehicles is another potential direction.

X. ACKNOWLEDGEMENT

The authors gratefully acknowledge the support of the Department of Computer Science and Information Technology, Dr. Babasaheb Ambedkar Marathwada University, Chhatrapati Sambhaji Nagar, Maharashtra, India.

REFERENCES

[1] Shankar Ka, Madhavan Va, Muhammad Tariq P Ra, and Sanjay Subramanian Ka,"Emergency Vehicle Detection in Heavy Traffic using Deep ConvNet2D and Computer Vision", Proceedings of the International Conference on Innovative Computing & Communication (ICICC) 2022.

[2] Pablo Barcellosa, Vitor Gomesa, and Jacob Scharcanskia, "Shadow Detection in Camera-based Vehicle Detection: Survey and Analysis ", Journal of Electronic Imaging, May 2016.

[3] Ana Filipa Rodrigues Nogueira, Hugo S. Oliveira, José J. M. Machado and João Manuel R. S. Tavares, " Sound Classification and Processing of Urban Environments: A Systematic Literature Review", Sensors 2022.

[4] Letizia Marchegiani and Ingmar Posner, "Leveraging the Urban Soundscape: Auditory Perception for Smart Vehicles", IEEE International Conference on Robotics and Automation (ICRA) Singapore, May 29 - June 3, 2017.

[5] Reagan L. Galvez, Argel A. Bandala, Elmer P. Dadios, Ryan Rhay P. Vicerra, and Jose Martin Z. Maningo, "Object Detection Using Convolutional Neural Networks", TENCON 2018 - 2018 IEEE Region 10 Conference, Jeju, Korea (South), 2018, pp. 2023-2027, doi: 10.1109/TENCON.2018.8650517.

[6] Ayoub Benali Amjoud and Mustapha Amrouch, "Object Detection Using Deep Learning, CNNs and Vision Transformers: A Review", IEEE, 2023.

[7] Justin Salamon and Juan Pablo Bello, "Deep Convolutional Neural Networks and Data Augmentation for Environmental Sound Classification", IEEE Signal Processing Letters, 2016.

[8] Jie Xie, Kai Hu, Mingying Zhu, Jinghu Yu, and Qibing Zhu, "Investigation of Different CNN-Based Models for Improved Bird Sound Classification ", IEEE, 2019.

[9] Ross Girshick, "Fast R-CNN", IEEE International Conference on Computer Vision, 2015.

[10] Sophiya E., and Jothilakshmi S., "Audio event detection using deep learning model", Int. J. Computer Aided Engineering and Technology, Vol. 16, No. 3, 2022.

[11] Hari Vignesh K, Musharraf U., and Balaji A,"Emergency Vehicle Detection using Deep Learning", IJARIIE-ISSN(O)-2395-4396, Vol-9 Issue-2, 2023.

[12] Dean Carmel, Ariel Yeshurun, and Yair Moshe, "Detection of Alarm Sounds in Noisy Environments", 25th European Signal Processing Conference (EUSIPCO), 2017.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 02, NCRTCS - 2026

[13] Paul Potnuru , Richard H. Epstein , Richard McNeer , and Christopher Bennett, "Development and Validation of an Algorithm for the Identification of Audible Medical Alarms", Article in Cureus, November 2020.
[14] Aiesha Roshan, Manukrishnan, Sneha Mohandas, and Sreejith PR, "Traffic Density Control Using Deep Learning", IJRES, Vol 10 Issue 6, 2022.
[15] K Agrawal, M K Nigam, S Bhattacharya, and Sumathi G, "Ambulance detection using image processing and neural networks", Journal of Physics: Conference Series, 2021.
[16] Van-Thuan Tran, and Wei-Ho Tsai, " Audio-Vision Emergency Vehicle Detection", IEEE SENSORS JOURNAL, VOL. 21, NO. 24, DECEMBER 15, 2021.
[17] Mingyuan Xin, and Yong Wang, "Research on image classification model based on deep convolution neural network", J Image Video Proc. 2019.
[18] Shuvendu Roy, and Md. Sakif Rahman, "Emergency Vehicle Detection on Heavy Traffic Road from CCTV Footage Using Deep Convolutional Neural Network", International Conference on Electrical, Computer and Communication Engineering (ECCE), 7-9 February, 2019.
[19] Aishwarya W, Mrs. Sheelavathi A, Jhanani R S, Karthika D R, and Keerthana P, "Detection of Ambulance Using Computer Vision", IJRASET, Volume 10 Issue VI June 2022.
[20] Usha Mittal, and Priyanka Chawla, "Acoustic Based Emergency Vehicle Detection Using Ensemble of Deep Learning Models", Elsevier, 2023.
[21] Venkatesh Boddapati, Andrej Petef, Jim Rasmusson, and Lars Lundberg, "Classifying environmental sounds using image recognition networks", International Conference on Knowledge Based and Intelligent Information and Engineering, 2017.
[22] Karpis, and Ondrej, "System for Vehicles Classification and Emergency Vehicles Detection", IFAC Proceedings Volumes, 2012.