

Cross-Attention Based Multi-Sensor Fusion for Robust Object Detection in ADAS: YOLOv12 with Spatiotemporal Calibration and iHOA-Tuned Temporal Fusion Transformer

Raghunath Mallavarapu

GITAM School of Science, GITAM Deemed to be University

Kanaka Raju Pappala

kpappala@gitam.edu

GITAM School of Science, GITAM Deemed to be University

Sattibabu Bhumireddi

GITAM School of Science, GITAM Deemed to be University

G. Krishna Podagatlapalli

GITAM School of Science, GITAM Deemed to be University

Kavitha Chandu

GITAM School of Science, GITAM Deemed to be University

Durga Rao Tadisetti

GITAM School of Science, GITAM Deemed to be University

Chandra Sekhar Angani

GITAM School of Science, GITAM Deemed to be University

Research Article

Keywords: Advanced Driver Assistance Systems, Multi-Sensor Fusion, Spatiotemporal Calibration, Cross-Attention Fusion, YOLOv12, Temporal Fusion Transformer

Posted Date: July 17th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10135180/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Cross-Attention Based Multi-Sensor Fusion for Robust Object Detection in ADAS: YOLOv12 with Spatiotemporal Calibration and iHOA-Tuned Temporal Fusion Transformer

Authors

Raghunath Mallavarapu¹, P. Kanaka Raju^{2*}, Bhumireddi Sattibabu³, G. Krishna Podagatlapalli⁴, C. Kavitha⁵, T. Durga Rao⁶, Chandra S. Angani⁷

Affiliations

¹Research Scholar, Department of Physics, GITAM School of Science, GITAM Deemed to be University, Visakhapatnam, Andhra Pradesh, India 530045.

ORCID: 0009-0002-8622-0647

Email: rmallava@gitam.in

²Department of Physics, GITAM School of Science, GITAM Deemed to be University, Visakhapatnam, Andhra Pradesh, India 530045.

ORCID: 0000-0003-1428-8992

Email: kpappala@gitam.edu

³Department of Physics, GITAM School of Science, GITAM Deemed to be University, Visakhapatnam, Andhra Pradesh, India 530045.

ORCID: 0000-0002-7405-9545

Email: sbhumire@gitam.edu

⁴Department of Physics, GITAM School of Science, GITAM Deemed to be University, Visakhapatnam, Andhra Pradesh, India 530045.

ORCID: 0000-0002-3644-8734

Email: gpodagat@gitam.edu

⁵Department of Physics, GITAM School of Science, GITAM Deemed to be University, Visakhapatnam, Andhra Pradesh, India 530045.

ORCID: 0000-0003-1516-924X

Email: kchandu@gitam.edu

⁶Department of Physics, GITAM School of Science, GITAM Deemed to be University, Visakhapatnam, Andhra Pradesh, India 530045.

ORCID: 0000-0001-6996-2919

Email: dtadiseta@gitam.edu

⁷Department of Physics, GITAM School of Science, GITAM Deemed to be University, Visakhapatnam, Andhra Pradesh, India 530045.

ORCID: 0000-0001-8728-4250

Email: cangani@gitam.edu

***Corresponding Author**

P. Kanaka Raju

Department of Physics, GITAM School of Science, GITAM Deemed to be University, Visakhapatnam, Andhra Pradesh, India 530045.

ORCID: 0000-0003-1428-8992

Email: kpappala@gitam.edu

Cross-Attention Based Multi-Sensor Fusion for Robust Object Detection in ADAS: YOLOv12 with Spatiotemporal Calibration and iHOA-Tuned Temporal Fusion Transformer

Abstract

Advanced Driver Assistance Systems (ADAS) have attracted massive interest recently, fueled by rapid breakthroughs in intelligent transportation and autonomous driving technologies. Accurately detecting surrounding objects is a critical requirement for ADAS, specifically under complex traffic environments where individual sensors frequently suffer from occlusion, noise, and limited sensing capability. To address these challenges, this paper introduces a Cross-Attention based Multi-Sensor Fusion framework for Robust Object Detection in Advanced Driver Assistance Systems (CAMSF-ADAS). The proposed model utilizes multimodal sensor data acquired from camera, LiDAR, and radar for robust perception and object recognition. Spatiotemporal calibration is performed to align camera, LiDAR, and radar data into a unified coordinate framework for enhancing spatial consistency and temporal synchronization among sensors. Besides, a Multi-Head Cross-Attention Fusion Module is introduced to efficiently integrate complementary information from different sensing modalities through dynamic feature interaction. For object detection, the YOLOv12 network is used to accurately localize and identify surrounding traffic objects under diverse driving scenarios. In addition, a Temporal Fusion Transformer (TFT) model is employed for classification to capture temporal dependencies and improve discriminative learning performance. To further improve the classification capability, the hyperparameters of the TFT model are optimized using the Improved Hippopotamus Optimization Algorithm, which enhances convergence behavior and search efficiency. Experimental analysis is carried out, and the obtained results show that the proposed CAMSF-ADAS model obtains superior performance compared to existing approaches. Therefore, the proposed model can be considered an effective solution for robust multi-sensor object detection in ADAS environments.

Keywords: Advanced Driver Assistance Systems; Multi-Sensor Fusion; Spatiotemporal Calibration; Cross-Attention Fusion; YOLOv12; Temporal Fusion Transformer

1. Introduction

The Advanced Driver Assistant Systems (ADAS) play a significant part in improving the road vehicles' safety. ADAS support drivers in performing complex driving tasks to prevent dangerous conditions [1]. Assisting studies contain: the activation of safety devices to mitigate imminent collisions, warning messages in dangerous driving situations, attention-less driver warnings, and autonomous maneuvers to avoid obstacles. Efforts to decrease the regularity and strictness of traffic accidents are intensely encouraged by both the legislative body and the automotive industries. The consistency of these methods is critical from both the safety perspectives and user acceptance. Individual drivers may become overly dependent on the ADAS model and fail to notice when they fail. Fake alarms and activities (like unnecessary emergency brake) typically result from failure of the situation perception mechanisms [2]. Although incredible progress has been made, problem identification remains a challenging task because driving situations are dynamic and generally include low illumination, occlusions, dense populations, heterogeneous objects, and adverse weather conditions like rain, snow, and fog [3]. Various ADAS mechanisms and autonomous vehicles rely on vision-based algorithms, namely image classification, to perceive the neighboring environment, such as the pedestrians, detection of lanes, obstacles, traffic signals, and road signs [4]. The developments of Deep learning (DL) have transformed the domain of ADAS by permitting strong tools for scene understanding, decision-making, and object recognition. To leverage progressive DL methodologies, the system effectively extracts complicated driving patterns and identifies anomalous or dangerous driving situations with improved accuracy [5].

1.1. Challenges in ADAS

Sensor fusion plays an important part in progressing the dependability of the perception method by merging data from different sensors and exploiting previous learning from diverse modalities. The sensor fusion technique, obtaining information from several different sensors to improve perception accuracy, fault-tolerance, and reliability, has gained increased attention as an outcome of these problems [6]. Radar, light detection and ranging (LiDAR), and vision-based (camera) sensors individually produce something exclusive to the perception system. Radar is good at measuring long-range speed and works reliably in all weather conditions. LiDAR gives very accurate 3D spatial details [7]. Vision systems help understand context and meaning from images. But each one has weaknesses: radar is not very precise in direction, LiDAR can struggle in dust or rain, and cameras are affected by lighting changes. Stronger obstacle sensing methods that significantly perform uni-modal resolutions are being produced through study by integrating these cross-modal sensing methodologies [8].

Various ADAS mechanisms and autonomous vehicles rely on vision-based algorithms, namely image classification, to perceive the neighboring environment, such as the pedestrians, detection of lanes, obstacles, traffic signals, and road signs [9]. However, as autonomous driving systems become more complex, particularly with the incorporation of different sensor modalities, the decision-making processes shown by Machine Learning methodology frequently led to a substantial absence of transparency [10]. Although these ML approaches are extremely effective at generalizing over a broad range of driving scenarios and are known to have a strong capability to method risk patterns by sophisticated data representation, their inner mechanisms and underlying decision-making process frequently lack in interpretability [11].

1.2. Motivation of the Study

Although several studies have been conducted on sensor fusion in ADAS, consistent and reliable object detection remains complex in practical driving settings. Multiple sensors, including LiDAR, radar, and camera, produce heterogeneous information, which creates difficulty in synchronization and alignment within multimodal learning. In various current techniques, sensor fusion is achieved by static fusion or simple weighted methodology, which may not effectively use the complementary data from multiple sensing modalities. Moreover, more recent approaches mostly focus on improving detection accuracy; limited attention is provided to temporal data learning and optimization of model hyperparameters. This can affect the recognition and categorization results in dynamic traffic situations. Thus, there is a necessity for an effective multi-sensor model with suitable preprocessing, fusion, sensor calibration, object detection, and classification for enhancing perception outcome in ADAS functions.

1.3. Proposed Contributions

Based on the above-discussed limitations, this work proposes a Cross-Attention-based Multi-Sensor Fusion framework for Robust Object Detection in Advanced Driver Assistance Systems (CAMSF-ADAS). The significant contributions of this study are as follows:

- Utilize a spatiotemporal calibration strategy to align camera, LiDAR, and radar data into a unified coordinate framework.
- Introduce a Multi-Head Cross-Attention Fusion Module (MHCAFm) for integrating complementary information from multiple sensing modalities.
- Employ the YOLOv12 object detection network for accurate localization and detection of surrounding traffic objects.
- Use a Temporal Fusion Transformer (TFT) for object classification by modeling temporal dependencies in sequential sensor data.
- Optimize the hyperparameters of the classification model using the Improved Hippopotamus Optimization Algorithm (IHOA) to improve learning performance.
- Perform extensive experiments on the benchmark dataset and highlight the improved performance of the model compared to existing approaches.

2. Related Works

In this section, existing related works on object detection in ADAS are presented along with the identified research gaps. Yang and Yi [12] presented a DL-based Intelligent Driver Drowsiness Detection for ADAS (DLID3-ADAS) approach. The presented method aims to improve road safety through the recognition of drowsiness among drivers. Falaschetti et al. [13] introduced a Tiny YOLO v3 and low-rank (LR) framework that meets the needs of real-time VaPD on an embedded model. The system has been established starting from Tiny YOLO v3, implementing a convolutional neural network (CNN) methodology according to Tucker tensor decomposition, capable of decreasing the computation complexity of the network. Gomes and Tamariz [14] examined the growth and authorization of a combined Advanced Driver-Assistance System (ADAS) merging YOLOv8-based computer vision with vehicle control in the CARLA simulator. The main goal was to perform and verify the method's ability

to identify stop signs and vehicles in the real world, delivering pictorial feedback to the driver and permitting the vehicle to respond to stop signs via stopping behavior and proper deceleration.

Murthy et al. [15] proposed a new approach algorithm, "A Real-Time Object Detection (OD) Framework for ADAS," by employing the advanced OD technique YOLOv5 that enhances the speed in detection of objects across the real world. This architecture is utilized to size a mobile application named "ObjectDetect" that supports individuals to make the best decisions on the road. Hong et al. [16] discussed a capacitance defect recognition method based on charge quantity (Q) examination. A dedicated key controller board was designed, incorporating analog-to-digital conversion and noise filtering using a graphical user interface (GUI). Sivakumar and Tripathi [17] evaluated and proposed DL-based OD approaches for real-world object identification. The presented model is utilized to represent its potential to improve the consistency of ADAS methods without the utilization of an image pre-processing module.

Al Amin et al. [18] addressed a novel, highly efficient Field Programmable Gate Array (FPGA), the basis of a real-world OD and classification method applying YOLO-v3 Tiny for edge computers. Yet, the presented technique has been symbolized with ADAS for assessment. Shahzad et al. [19] investigated this problem by presenting a new methodology: applying DDNN as a preprocessing stage to transmute adverse weather images with clear weather images, thus improving the strength of the ML-ADAS approach. An overview of existing related works on OD in ADAS is presented in Table 1.

Table 1 Summary of existing research on ADAS

Author/Year	Method	Dataset	Objective	Accuracy
Yang et al. (2024) [12]	Deep Learning Model	YawDD driver dataset	To enhance road safety with a driver drowsiness detection system	97.05% Accuracy
Falaschetti et al. (2024) [13]	low-rank (LR) Tiny YOLO v3 method	PASCAL VOC and KITTI datasets.	Real-time vehicle and pedestrian monitoring through embedded systems	Attains better accuracy
Gomes & Tamariz (2026) [14]	YOLO-based Driver Assistance model	CARLA simulator dataset	To develop real-time detection for stop signs and vehicles in an autonomous driving system.	High real-time detection accuracy
Murthy et al. (2022) [15]	YOLOv5 model	Three benchmark datasets	To enhance live OD speed and tracking in an ADAS environment	95 % Accuracy
Hong et al. (2026) [16]	capacitance defect detection algorithm	Inspection of over 240,000 capacitive touch pads	To improve quality assurance and safety in ADAS	Achieved reliable real-time detection

Sivakumar et al. (2025) [17]	CNN and DNN models	CIFAR-10, CIFAR-100, and YOLO datasets	To enhance real-time detection for safer ADAS	Achieved better detection accuracy and speed
Al Amin et al. (2024) [18]	FPGA and YOLO-v3 tiny model		To develop low-power real-time traffic-light detection and classification for ADAS	99% accuracy at 15 FPS with 3.5W power consumption
Shahzad et al. (2024) [19]	Weather UNet DNN model	KITTI dataset	To improve OD performance in challenging environmental settings	Achieved robust detection in extreme fog conditions

2.1. Limitations of Existing Methods

Yang and Yi [12] proposed a drowsiness detection method utilizing DL and optimizer approaches, then the concentration is mostly restricted to driver state monitoring rather than a multi-sensor environmental perspective. Falaschetti et al. [13] presented a lightweight YOLO-driven method for real-life identification, but the model is exploited for single-sensor input and doesn't solve the multi-modal fusion task. Murthy et al. [15] introduced a YOLOv5-driven real-world object detection for applications of ADAS, because the method doesn't consider temporal dependency or sophisticated fusion approaches. Likewise, Shahzad et al. [19] improved ADAS robustness with adverse weather states, but the approach is based on pre-processing improvement. It doesn't fully combine multi-sensor calibration and attention-driven fusion modules. In general, most existing models either focus on unique tasks like classification or detection, or they utilize a restricted fusion approach that fails to fully use complementary data from diverse sensors in intricate driving systems.

3. Proposed Methodology

This section describes the overall workflow of the proposed CAMSF-ADAS approach and its key components.

3.1. Overall Framework of Proposed Model

The overall process of the proposed model is illustrated in Fig. 1. The input data is first gathered from numerous sensors, including camera, LiDAR, and radar. These raw inputs are then passed through a preprocessing stage where robust scaling is used to standardize the data and minimize the effect of extreme values. After preprocessing, a spatiotemporal calibration module is used to align the sensor data in both space and time so that all modalities follow a common reference frame. The aligned features are then fed to the fusion module, where a MHCAFm is utilized to integrate information from different sensors. The fused feature representation is then passed to the YOLOv12 network for object detection, where surrounding objects are identified and localized. The detected outputs are further processed using a TFT to learn sequential dependencies and enhance classification performance. Eventually, an IHOA is applied to tune the model parameters and improve the overall performance of the system.

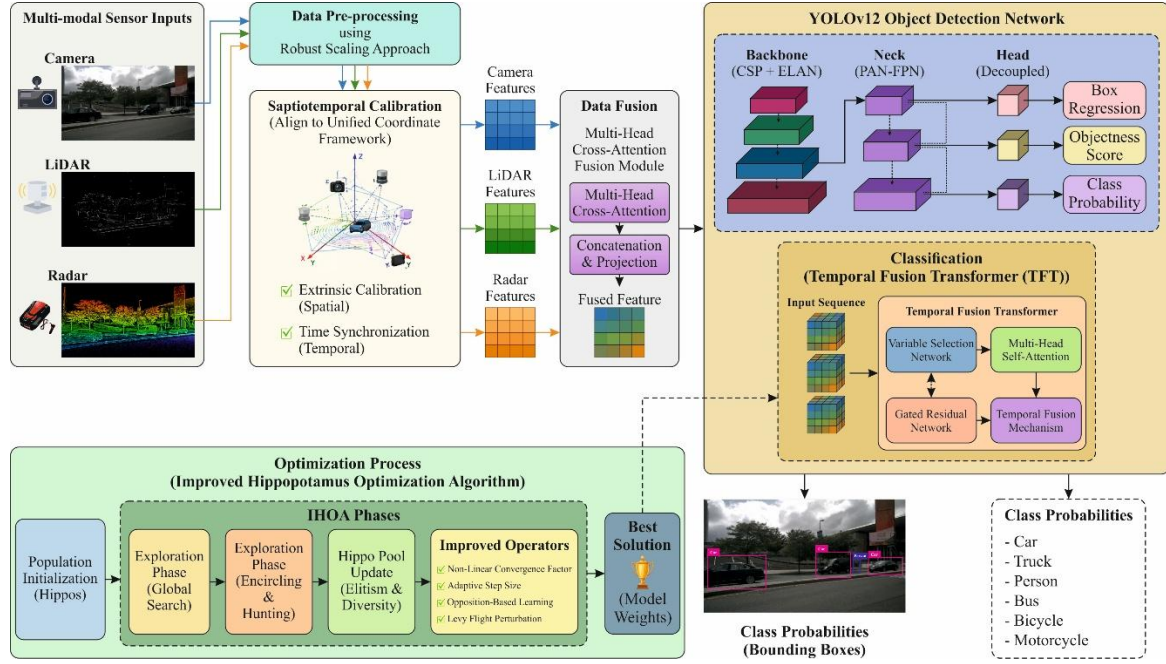


Fig. 1. Overall architecture of the proposed model

3.2. Robust Scaling-Based Data Preprocessing

This subsection elucidates the preprocessing step used for normalizing camera, LiDAR, and radar sensor data prior to the spatiotemporal calibration stage. Robust scaling minimizes the influence of outliers through scaling data using the median and IQR, making it suitable for extreme values [20]. It is applied when the dataset consists of many outliers, and it is essential to preserve the relative distance among non-outlier data points. It subtracts the dataset median and divides the result by the IQR, which minimizes the impact of outliers and retains the distribution of non-outlier data values. The median and the IQR are preserved in the fitting process so that they can be utilized for future data through the transform method.

$$X_{\text{scaled}} = \frac{X - \text{Median}(X)}{Q_3 - Q_1} \quad (1)$$

Here, the original value is denoted by, the first quartile is represented as Q_1 , the term $\text{Median}(X)$ indicates the median quartile or Q_2 , and the third quartile is denoted as Q_3 .

3.3. Spatiotemporal Sensor Calibration

This subsection explains the calibration process used to align camera, LiDAR, and radar data in a common spatial and temporal reference frame [21].

Target-based techniques: These approaches determine the cross-sensor data correlation by presenting exactly developed calibration objects. General examples comprise AprilTags or checkerboards for geometric feature boards for LiDAR, cameras, and corner reflector arrays in the radar.

Targetless methods: These approaches learn features directly from natural settings like SURF or SIFT keypoints, planar surfaces, or lines, and utilize the cross-sensor feature constancy restraints to form the data correlation.

Hence, their performance mainly relies on the feature counts in natural settings. In the spatiotemporal extrinsic calibration, the conversion amongst various sensors can be denoted by,

$$P_t = RP_s + T \quad (2)$$

In this mathematical form, the parameter R and T remain the rotation matrix and the translation vector, the P_s refers to the source sensor co-ordinates, and the term P_t characterizes the converted target sensor operators. With regard to temporal alignments, the calculation of asynchronous sensors has been synchronized by employing the timestamp matching:

$$\Delta t = t_r - t_c \quad (3)$$

Here, the term t_r signifies the timestamp of LiDAR or radar data, and t_c indicates the timestamp of camera data.

Particularly, the camera–LiDAR–radar calibration approaches are combined with the learning-based systems, information-based approaches, and feature-based techniques. In this work, targetless spatiotemporal measurement has been utilized to balance the radar camera and LiDAR data in the integrated coordinate system before the MHCAFM.

3.4. Multi-Head Cross-Attention Fusion Module

This subsection defines the proposed fusion mechanism used to integrate features from multiple sensors using cross-attention. After obtaining input from camera, LiDAR, and radar data and combine them effectively. The data obtained from these modalities are complementary [22]. Camera sensors capture detailed texture and visual information, LiDAR sensors offer spatial depth information, and radar sensors provide the measurement relevant to object motion and distance. This framework utilizes the complementary nature of heterogeneous sensor modalities through exchanging their feature information to achieve effective fusion. Initially, cross-attention (CA) is used to exchange complementary information among various modalities. The processes of CA are specified: For the input feature maps of camera $F_c \in R^{H \times W \times C}$, LiDAR $F_l \in R^{H \times W \times C}$, and radar $F_r \in R^{H \times W \times C}$, the feature normalization is performed in the initial stage, the mapping features are then reshaped and linearly changed to produce query (Q), key (K), and value (V).

$$F_c = \text{reshape}(LN(F_c)), F_l = \text{reshape}(LN(F_l)), F_r = \text{reshape}(LN(F_r)) \quad (4)$$

$$Q = F_c W_Q, K = F_l W_K, V = F_r W_V \quad (5)$$

Here, LayerNorm is denoted by LN (), and the terms W_Q , W_K , and W_V represent the linear projection matrices. The cross-attention approach evaluates the similarity among the query and the key. The outputs are weighted to attain the attention scores that supports the exchange of context among the sensor modalities.

$$\text{head}^i = CA(Q^i, K^i, V^i) \quad (6)$$

$$MCA(F_c, F_l, F_r) = \text{concat}(\text{head}^1, \dots, \text{head}^i, \dots, \text{head}^n) \quad (7)$$

Here, the term MCA indicates the feature outcome of the cross-attention module, and head^i denotes the outcome of the i_{th} head. The multi-model feature representation map is formed by concatenating the interaction feature along with the dimension of channel.

$$F_m = C(MCA(F_c, F_l, F_r)) \quad (8)$$

where $C()$ indicates the concatenation process along the channel dimension. At the final stage, the combined features were passed through LayerNorm and a nonlinear mapping to achieve the final fused mapping feature F_{fuse} . Moreover, a residual connection is incorporated to reduce the loss of information.

$$F_{fuse} = LN(F_m) + F_m \quad (9)$$

In this study, the MHCAFm is utilized to effectively combine camera, LiDAR, and radar features to strengthen multimodal feature representation before object detection. Fig. 2 indicates the multi-head cross-attention fusion of camera, LiDAR, and radar features.

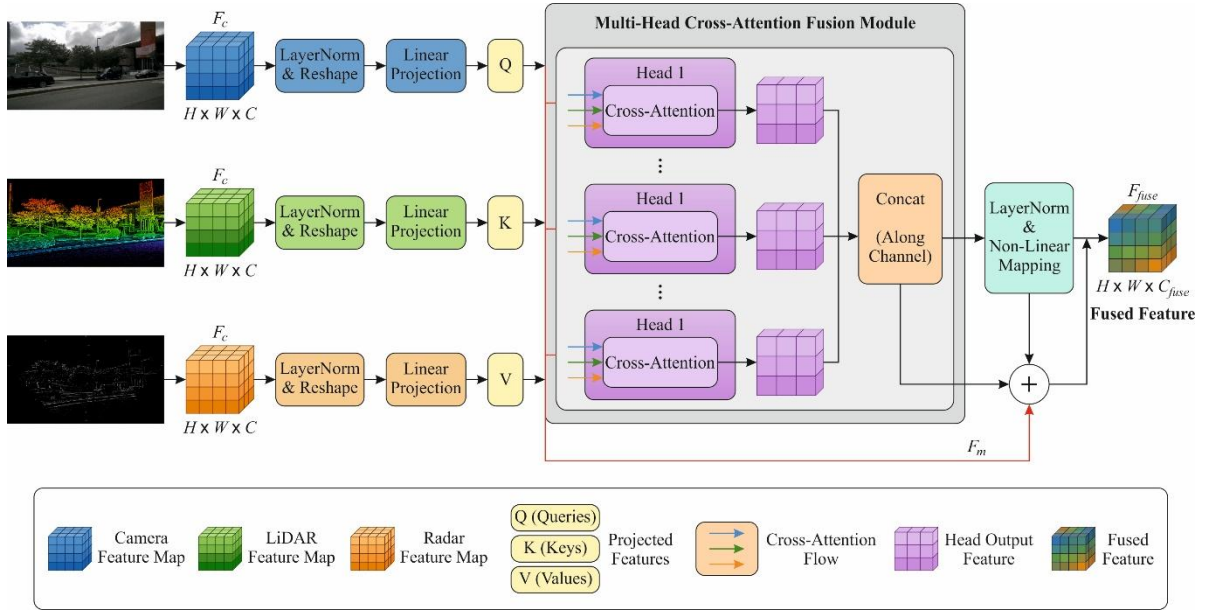


Fig. 2. Multi-head cross-attention fusion of camera, LiDAR, and radar features

3.5. YOLOv12 Object Detection Network

This subsection presents the YOLOv12-based detection model leveraged to localize objects in the driving environment. The YOLOv12 remains the current development in the YOLO series. Unlike prior types, the YOLOv12 presents architectural improvements and integrates recent technologies that result in prominent gains in the recognition performance and computation effectiveness [23]. This model provides five diverse forms: n, s, m, l, and x. All individuals are increasing detection accuracy and network depth to encounter the fluctuating requirements of various applications. Later, the comprehensive evaluation of factors like hardware compatibility, detection accuracy, and model complexity. Hence, the YOLOv12-n is the base framework for this analysis. The YOLOv12 model is built with three major portions: backbone, neck, and head. The backbone is responsible for extracting features from images. It uses an improved R-ELAN design that helps reuse features efficiently while keeping the model lightweight and stable during training. It also uses larger 7x7 convolutions to better understand complex traffic scenes. The neck combines features from different scales using an attention-based method. This helps the model concentrate on important regions in the image. It also improves speed and memory usage with FlashAttention, while still capturing useful context from multiple scales. The head makes the final predictions. It

separates classification and location tasks so they don't interfere with each other. It also uses lighter convolution layers to reduce computation and improve efficiency. Eventually, YOLOv12-n is designed to be fast, and efficient for making it well suited for object detection in ADAS applications.

3.6. Temporal Fusion Transformer for Classification

This subsection defines the TFT utilized for learning temporal relationships in the fused sensor data. TFT is a neural network (NN) framework that depends on the attention mechanism (AM) that integrates temporal and transformer modeling methods for handling sequential details with numerous features [24]. The major elements of TFT comprise temporal processing units, variable selection networks, a gated residual network (GRN), and static covariate encoders. Multifaceted networks in DL architecture cannot essentially realize superior predictive efficiency. The TFT employs GRN architecture for simplifying the model by disregarding redundant parts in the architecture, thus permitting the TFT module to perform well in additional difficult settings and handling enormous datasets. The input of GRN comprises dual vectors, such as the primary data p vector and the contextual data vector c . The major calculation method of GRN can be represented by,

$$GRN_W(p, c) = LlayerNorm(p + GLU_W(h_1)) \quad (10)$$

$$h_1 = W_{1,w}h_2 + b_1 \quad (11)$$

$$h_2 = ELU(W_{2,w}p + W_{3,w}c + b_W) \quad (12)$$

Now the parameters h_1 and h_2 signify the outcomes of dual intermediary layers. The ELU activation function can be derived by

$$f(x) = \begin{cases} x, & x > 0 \\ (e^x - 1), & x \leq 0 \end{cases} \quad (13)$$

The Gated Linear Unit (GLU) is a significant module that achieves the gating function. By using the GLU module, the GRN architecture is proficient in directing the model structures and disregarding redundant layers. Considering the input is denoted as g , the GLU can be determined by,

$$GLU_W(g) = s(W_{4,w}g + b_{4,w})e(W_{5,w}g + b_{5,w}) \quad (14)$$

Here, the term b represents the biases, the factors W signify weights, and the parameter s represents the sigmoid activation function.

The correlation of input and target variables can be generally challenged for evaluation. Then, variable selection networks have been presented in the TFT. The variable selection network automatically recognizes and chooses variables, which can be significant for the classification process, therefore avoiding irrelevant data. This adaptive variable selection technique allows the TFT to achieve greater efficiency and accuracy while handling higher-dimensional sequential details.

The TFT encode the input feature as vectors with d dimensions, here the sizes of the vectors are equal to the node counts across the network. The variable selection network weights can be determined by,

$$y_{c_t} = softmax(GRN_{V_{c_t}}(x_t^{(i)}, c_s)) \quad (15)$$

Here, the parameter c_s represents the context vector formed on the static covariate encoder, and the term x_t refers to the outcome of the gated residual network. The temporal processing model is the main component of TFT that employs the AM of the transformer to obtain the sequential sensor data. By employing the integration of the SAM, the component is capable of capturing dependencies amongst diverse positions in the sequences and producing feature representations that comprise extensive information.

The SAM can be computed as given below:

$$Attn(Q, K, V) = A(Q, K)V \quad (16)$$

Now the parameter V represents the value, K means the key, and Q refers to the query. The normalization operation can be determined by:

$$A(Q, K) = Softmax(QK^T\sqrt{d_{atn}}) \quad (17)$$

The TFT utilizes the multi-head attention (MHA) approach to increase the capability of model fitting, and the mathematical form is given as:

$$MultiHaed(Q, K, V) = [H_1, \dots, H_{m_H}]W_H \quad (18)$$

$$H_h = Attn(QW_o^{(h)}, KW_K^{(h)}, VW_y^{(h)}) \quad (19)$$

3.7. Improved Hippopotamus Optimization Algorithm

This subsection explains the optimization strategy used to tune the model hyperparameters and improve overall performance. The HOA is an innovative method for metaheuristic optimization. Based on the behavior of hippopotamuses while directing groups, protecting themselves against threats, and fleeing briskly, the method searches these activities to control the optimization procedures [25]. As a population-based technique, the HOA is efficient with respect to exploitation and exploration, thus making it a valuable approach for addressing difficult and multivariable optimization complexity.

Inspiration: The behavioral model of HOA depends upon the 3 key stages noticed in hippopotamuses:

Exploration stage: The HO update their location when interconnecting with the leading individual and the residue of the group.

Defense phase: As regards predator threats, the hippopotamus can change strongly and make an effort to prevent the invader.

Exploitation phase: When defense fails, the hippopotamuses retreat quickly towards the secure position.

These natural characters are arithmetically encoded to execute local and global search successfully at the time of the optimization method.

Initialization: Considered the search space is limited among the upper bound (ub) and lower bound (lb). The location of every individual hippopotamus has been modified by:

$$x_{i,j} = lb_j + r \cdot (ub_j - lb_j) \quad (20)$$

The population matrix can be indicated by,

$$X = \begin{bmatrix} x_{1,1} & \cdots & x_{1,m} \\ \vdots & \ddots & \vdots \\ x_{N,1} & \cdots & x_{N,m} \end{bmatrix} \quad (21)$$

Here, the term m signifies the search dimensions, and parameter N represents the population size. The improvement approaches presented in this study comprise two main modules. Primarily, the chaotic map initialization employs the logistic chaotic mapping to produce different initial populations, thus enhancing the solution space [26]. Secondly, the adaptive manipulation component dynamically modifies the strength of exploitation to better enable search over diverse iteration phases.

Chaotic Map Initialization Strategy and Its Role: According to the global search performance, the diversity of the early population is the major influence factor. It presents the population initialization technique that depends upon the logistic chaotic maps in the typical HOA to improve the variation of the early outcomes and increase the exploration proficiency in the earlier search phases.

$$X_{n+1} = r \times X_n \times (1 - X_n) \quad (22)$$

From this mathematical form, the term r indicates the chaotic factor, generally set to 4. The main outcomes produced by employing the chaotic mapping must consistently cover the hunting space and improve the global searching ability of the model.

Design and Implementation of the Adaptive Exploitation Mechanism: At the conventional HOA, the exploitation intensity can be constant. The count of iterations rises, and the search ability progressively drops. To handle this complexity, the model uses an adaptive approach where the level of exploitation changes dynamically as the iteration counts increases.

$$w = w_{\min} + (0.9 - w_{\min}) \times \left(1 - \frac{t}{t_{\max}}\right) \quad (23)$$

Here, the parameter $t_{\max}$ implies the overall iteration counts, and the factor t stands for the existing iteration.

Optimization Process

The proposed IHOA is utilized to fine-tune the hyperparameters of the TFT mechanism. The algorithm repeatedly updates the candidate solution to select an effective hyperparameter configuration to enhance the classification performance. The optimization process includes parameters like learning rate, batch size, epoch counts, and dropout rate. The fitness function is evaluated using classification accuracy, defined as follows:

$$Fitness = \max(Accuracy) \quad (24)$$

The main aim is to reduce the classification accuracy of the proposed model.

4. Experimental Validation

This section offers the experimental setup utilized for calculating the performance of the proposed model.

4.1. Dataset Description

This subsection provides details about the Nuscenes Dataset [27] and the CARRADA dataset [28] utilized for training and testing the proposed model. The following is a brief description of the datasets.

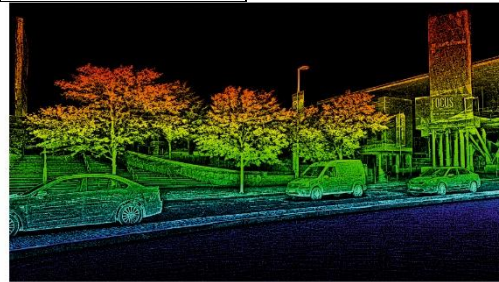
Nuscenes Dataset: The dataset is a large-scale autonomous driving dataset developed for computer vision and self-driving car studies. It comprises data collected utilizing six cameras, one LiDAR, and five radar sensors, offering a complete 360⁰ view of driving settings. It is broadly employed for tasks, namely object detection, tracking, sensor fusion, and autonomous vehicle perception modeling. As displayed in Table 2, the dataset encompasses six distinct classes, with every class consisting of 250 samples, resulting in a total of 1,500 samples overall. Fig. 3 depicts the samples of camera, LiDAR and Radar images. Fig. 4 illustrates the samples of predicted image.

Table 2 Details of the dataset

Classes	No. of Samples
“Car”	250
“Truck”	250
“Person”	250
“Bus”	250
“Bicycle”	250
“Motorcycle”	250
Total Samples	1500



Camera Data



LiDAR Data



Radar Data

Fig. 3. Samples of camera, LiDAR and Radar images

CARRADA Dataset: The CARRADA dataset encompasses synchronized camera and automotive radar recordings with 30 sequences of different scenarios with more than two moving objects. The radar data is processed using a semi-automatic approach. This open-source dataset provides RAD tensors along with detailed semantic annotations for both RD and RA views. The data is categorized into four classes: pedestrian, cyclist, car, and background. Each RAD tensor is provided with a size of $256 \times 256 \times 64$.

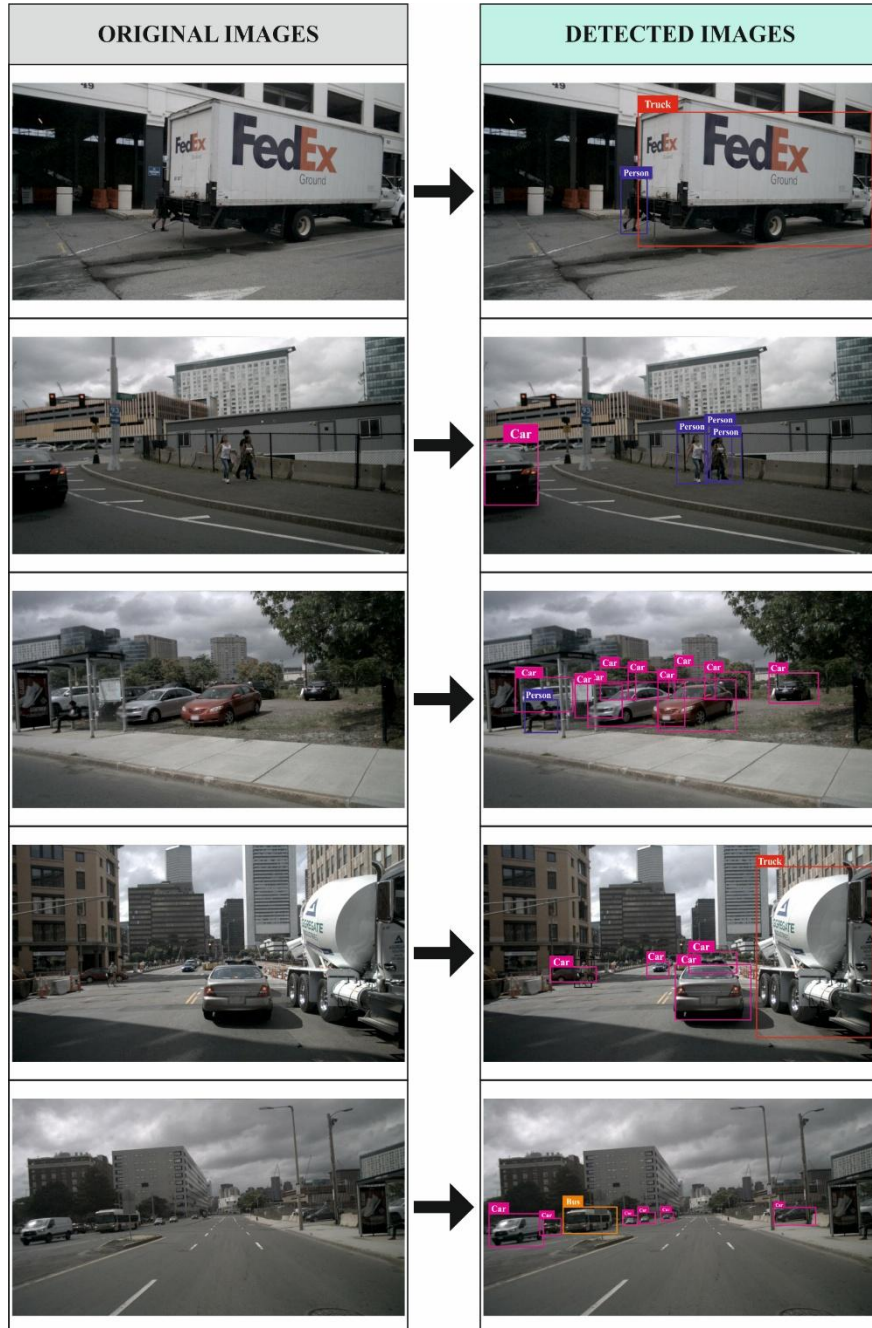


Fig. 4. Sample of predicted images

4.2. Evaluation Metrics

This subsection explains the metrics used to evaluate the proposed model. The study uses accuracy, precision, recall, F1-score, MCC, IoU, mIoU, Dice score, and mDice to measure both classification and segmentation performance. Here, TP, TN, FP, and FN denote true positives, true negatives, false positives, and false negatives, respectively.

$$Accuracy = (TP + TN) / (TP + TN + FP + FN) \quad (25)$$

$$Precision = TP / (TP + FP) \quad (26)$$

$$Recall = TP / (TP + FN) \quad (27)$$

$$F1 - score = (2 \times Precision \times Recall) / (Precision + Recall) \quad (28)$$

$$MCC = (TP \times TN - FP \times FN) / \sqrt{((TP + FP)(TP + FN)(TN + FP)(TN + FN))} \quad (29)$$

$$IoU = TP / (TP + FP + FN) \quad (30)$$

$$mIoU = (1 / N) \times \sum IoU_i \quad (31)$$

$$Dice = (2 \times TP) / (2 \times TP + FP + FN) \quad (32)$$

$$mDice = (1 / N) \times \sum Dice_i \quad (33)$$

Finally, computational time is also utilized for evaluating the efficacy of the model during inference.

5. Results and Discussion

This section discusses the experimental assessment and compares the proposed method with existing approaches.

5.1. Confusion Matrix Analysis

Fig. 5 illustrates the confusion matrices created by the CAMSF-ADAS algorithm for the (a-f) 500- 3000. The solution shows that the CAMSF-ADAS methodology precisely classifies and detects each of the six classes.

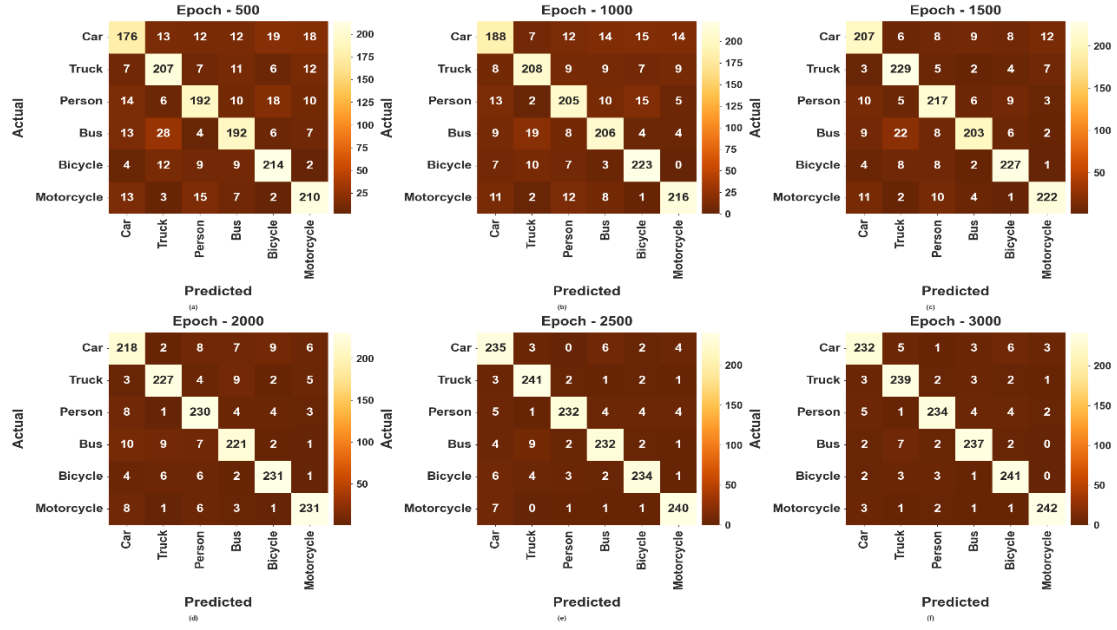


Fig. 5. Confusion matrices of the CAMSF-ADAS method with (a-f) Epoch 500-3000

Table 3 presents the object detection outcomes of the CAMSF-ADAS system with epoch 500-3000. The result implied that the CAMSF-ADAS approach accurately detected the image samples. Based on 500 epoch, the CAMSF-ADAS technique achieves result with $accuracy$ of 93.13%, $precision$ of 79.39%, $recall$ of 79.40%, $F1_{score}$ of 79.32%, and MCC of 75.26%, respectively. In addition, on 1500, the CAMSF-ADAS method obtained results with $accuracy$ of 95.67%, $precision$ of 87.09%, $recall$ of 87.00%, $F1_{score}$ of 86.97%, and MCC of 84.43%, respectively. Finally, on 3000, the CAMSF-ADAS system attained results with $accuracy$ of 98.33%, $precision$ of 95.02%, $recall$ of 95.00%, $F1_{score}$ of 95.00%, and MCC of 94.01%, respectively.

Table 3 Object detection result of the CAMSF-ADAS approach with Epoch 500-3000

Classes	Number of Epochs					
	500	1000	1500	2000	2500	3000
Accuracy	93.13	94.36	95.67	96.84	98.09	98.33
Precision	79.39	83.03	87.09	90.56	94.32	95.02
Recall	79.40	83.07	87.00	90.53	94.27	95.00
F1-Score	79.32	83.03	86.97	90.54	94.27	95.00
MCC	75.26	79.66	84.43	88.65	93.14	94.01

5.2. Performance Comparison with Existing Methods

Table 4 and Fig. 6 reveal the comparison study of the CAMSF-ADAS approach with other techniques using evaluation metrics [29–32]. The stimulation result of the CAMSF-ADAS system achieves better performance. The HRNetV2p-w18 method has gained lower value with $accuracy$ of 69.21%, $precision$ of 62.40%, $recall$ of 77.92%, and $F1_{score}$ of 67.81%, while the CRF-Net, HRFuser-T, and MV3D approaches have got slightly higher outcomes. Also, the NVLidarNet, SHJU-HW, and PIXOR++ models have attained moderate performance. At the same time, the CAMSF-ADAS technique has achieved greater outcome with $accuracy$ of 98.33%, $precision$ of 95.02%, $recall$ of 95.00%, and $F1_{score}$ of 95.00%, respectively.

Table 4 Comparative results of the CAMSF-ADAS method with existing systems

Method	$Accur_y$	$Preci_n$	$Recal_l$	$F1_{score}$	Computational Time (sec)
HRNetV2p-w18	69.21	62.40	77.92	67.81	4.93
CRF-Net	89.64	67.00	72.17	63.04	5.21
HRFuser-T	71.22	64.60	67.06	63.01	7.90
MV3D	70.63	66.77	93.47	89.70	5.61
NVLidarNet	90.27	91.70	87.32	90.88	4.37
SHJU-HW	93.85	94.14	85.79	92.78	7.40
PIXOR++	90.38	89.27	90.69	88.97	5.12
CAMSF-ADAS	98.33	95.02	95.00	95.00	3.28

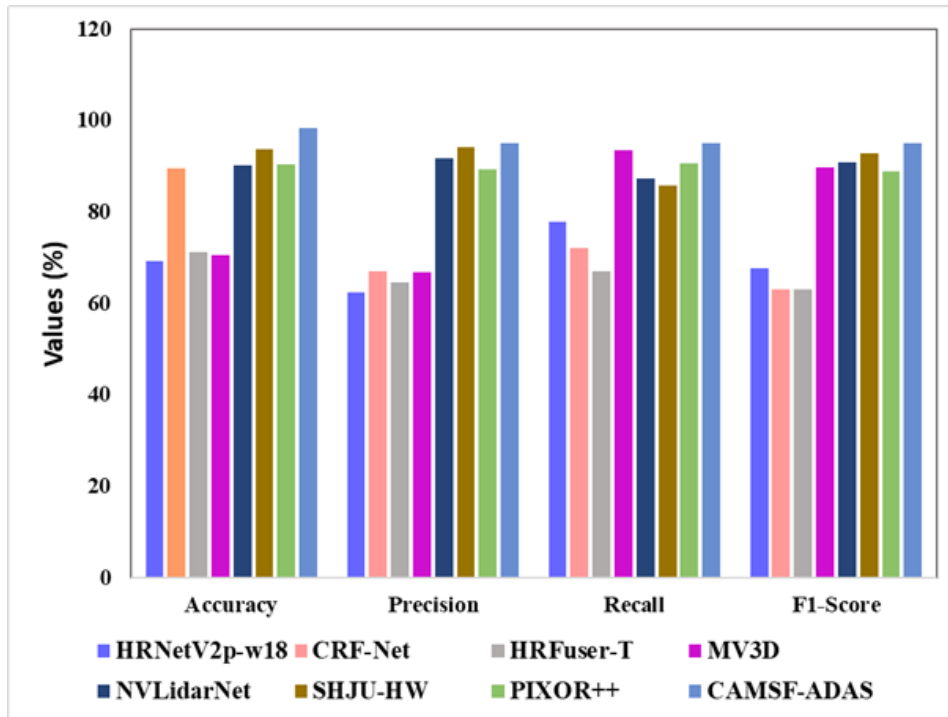
**Fig. 6.** Comparison result of the CAMSF-ADAS approach with various metrics

Fig. 7 displays the computational time(CT) of the CAMSF-ADAS system with other systems. Based on CT, the CAMSF-ADAS approach has reached minimal performance with CT of 3.28sec, whereas the HRNetV2p-w18, CRF-Net, HRFuser-T, MV3D, NVLidarNet, SHJU-HW, and PIXOR++ methodologies have gained maximal performance with CT of 4.93sec, 5.21sec, 7.90sec, 5.61sec, 4.37sec, 7.40sec, and 5.12sec, respectively.

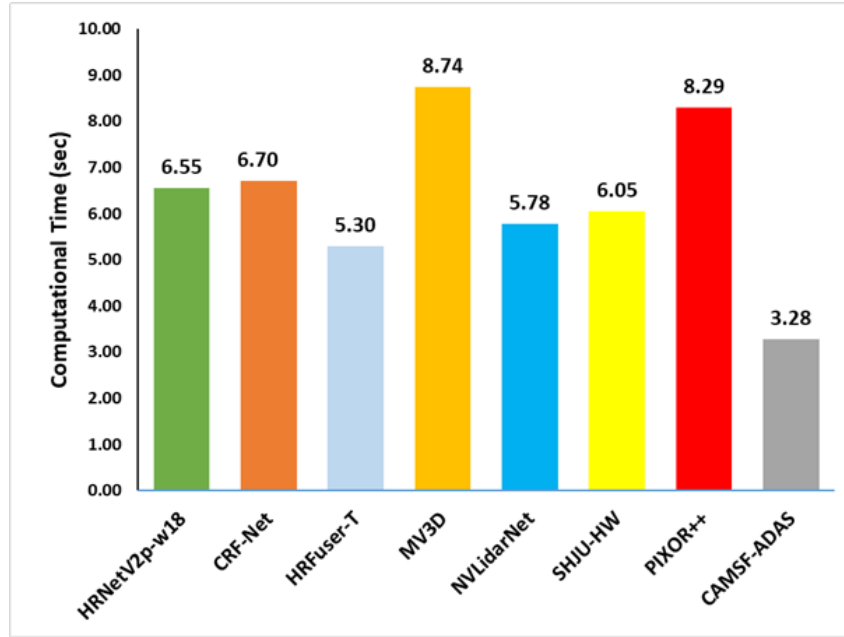


Fig. 7. Computational time of the CAMSF-ADAS approach with existing systems

Table 5 and Fig. 8 depict a comparative IoU analysis of the CAMSF-ADAS model against existing models under various classes. The outcome represents that the CAMSF-ADAS model has obtained greater outcomes in Background of 99.94%, Pedestrian of 33.39%, Cyclist of 20.79%, Car of 48.66%, and mIoU of 50.70%, whereas the existing models, F-PointNet, AVOD-FPN, Radar-Camera Fusion, Cont-Fuse, HDNET++, VoxelNet, and MMF approaches got lower outcomes.

Table 5 Comparison of IoU Analysis for CAMSF-ADAS model

IoU (%)					
Technique	Background	Pedestrian	Cyclist	Car	mIoU
F-PointNet	99.82	4.89	7.43	23.04	33.80
AVOD-FPN	99.44	8.69	6.98	17.11	33.06
Radar-Camera Fusion	99.71	2.88	3.90	8.26	28.69
Cont-Fuse	99.74	27.29	9.85	32.02	42.23
HDNET++	99.84	26.66	10.52	40.50	44.38
VoxelNet	99.83	26.45	12.46	40.65	44.85
MMF	99.82	30.13	12.06	42.91	46.23
CAMSF-ADAS	99.94	33.39	20.79	48.66	50.70

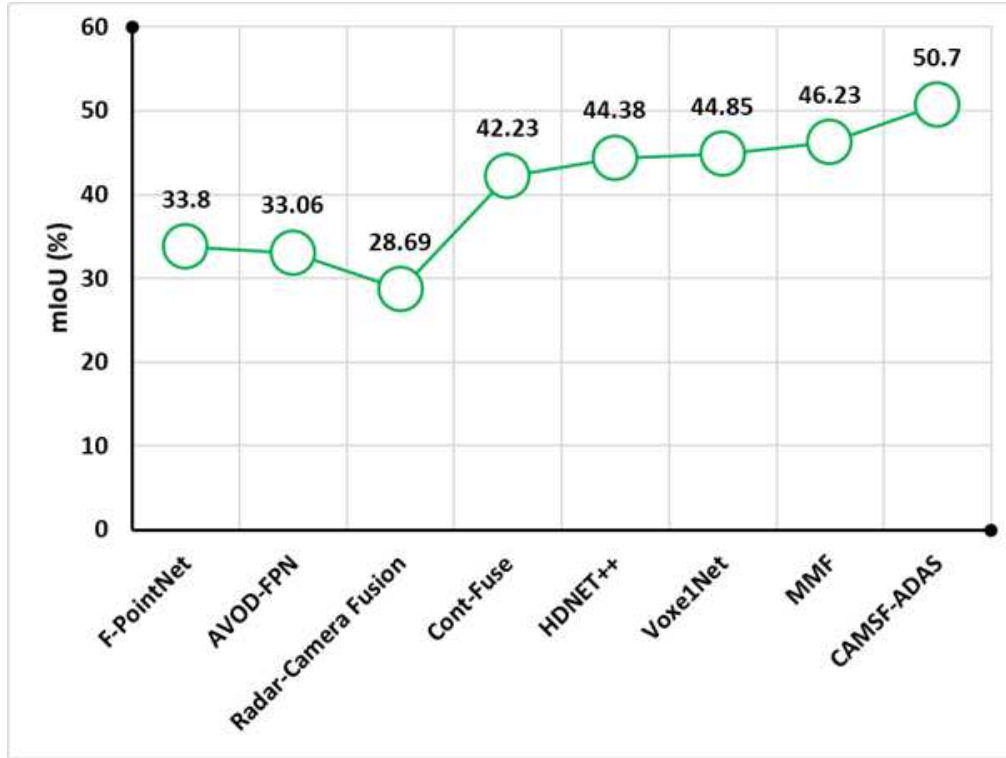


Fig. 8. Comparison of IoU analysis for CAMSF-ADAS model

Table 6 and Fig. 9 signify a comparative Dice analysis of the CAMSF-ADAS model against existing models across multiple classes. The outcome shows that the F-PointNet, AVOD-FPN, and Radar-Camera Fusion models had ineffective performance. Moreover, the Cont-Fuse, HDNET++, VoxelNet, and MMF techniques attain slightly lesser performance. In the meantime, the proposed CAMSF-ADAS model gains superior performance in Background of 99.91%, Pedestrian of 51.98%, Cyclist of 31.57%, Car of 62.91%, and mIoU of 61.59%.

Table 6 Comparison of Dice analysis

Dice (%)					
Technique	Background	Pedestrian	Cyclist	Car	mDice
F-PointNet	99.81	7.70	12.44	36.85	39.20
AVOD-FPN	99.72	15.06	11.87	28.61	38.82
Radar-Camera Fusion	99.85	4.74	6.20	14.78	31.39
Cont-Fuse	99.82	42.76	17.25	48.32	52.04
HDNET++	99.85	41.69	18.64	57.87	54.51
VoxelNet	99.86	41.48	21.76	57.98	55.27
MMF	99.84	45.52	20.89	59.79	56.51
CAMSF-ADAS	99.91	51.98	31.57	62.91	61.59

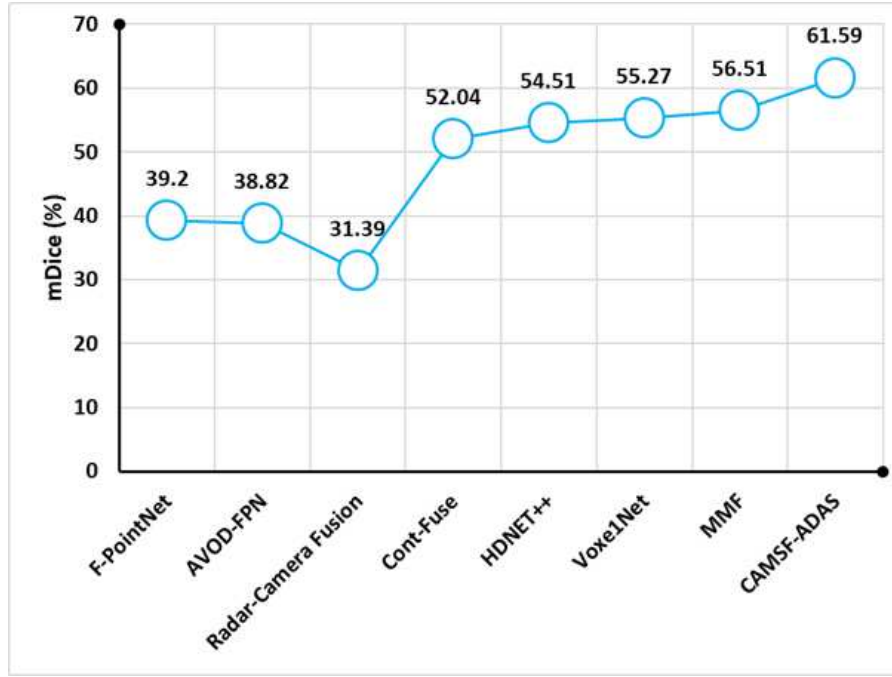


Fig. 9. Comparison of Dice analysis for CAMSF-ADAS model

5.3. Ablation Analysis

An ablation study examines the significance of various model elements by eliminating or changing them and examining the performance variations. It supports classifying which parts contribute most to the model's efficiency. This procedure helps enhance technique understanding and optimization. Table 7 denotes the ablation study of the CAMSF-ADAS methodology. The outcome demonstrated that the YOLOv12 only, TFT only, YOLOv12 + LSTM, YOLOv12 + BiLSTM, YOLOv12 + Transformer, YOLOv12 + TFT with Adam, and YOLOv12 + TFT (without IHOA) approaches got minimal performance. At the same time, the CAMSF-ADAS (YOLOv12 + TFT + IHOA) model gains maximal performance with a $accu_r$ of 98.33%, $preci_n$ of 95.02%, $recal_l$ of 95.00% and $F1_{score}$ of 95.00%.

Table 7 Ablation study of the CAMSF-ADAS method with metrics

Method	$Accu_r$	$Preci_n$	$Recal_l$	$F1_{score}$
YOLOv12 only	93.69	90.77	90.30	90.24
TFT only	94.43	91.47	91.00	90.95
YOLOv12 + LSTM	95.18	92.05	91.76	91.55
YOLOv12 + BiLSTM	95.81	92.58	92.49	92.07
YOLOv12 + Transformer	96.43	93.19	93.06	92.87
YOLOv12 + TFT with Adam	97.07	93.71	93.77	93.63
YOLOv12 + TFT (without IHOA)	97.73	94.28	94.27	94.36
CAMSF-ADAS (YOLOv12 + TFT + IHOA)	98.33	95.02	95.00	95.00

5.4. Discussion

On the main detection benchmark, CAMSF-ADAS attains the highest accuracy (98.33%) in comparison with other approaches, namely, CRF-Net, MV3D, and HRNetV2p-w18. It preserves a good balance among recall, F1-score, and precision, which displays that the method is not only precise but also consistent in identifying every object class. During the CARRADA dataset experimentations, the proposed technique again exposes the best outcome in both Dice scores and IoU. For each object type, such as cyclist, car, and pedestrian, CAMSF-ADAS commits to high segmentation quality associated with models like MMF and Cont-Fuse. This shows that the fusion module is capable of capturing complementary data from camera and radar information. In the ablation study, every component evidently contributes to the final results. TFT alone and YOLOv12 alone contribute low performance, whereas increasing sequence modeling and fusion progressively enhances outcomes. The last integration with TFT, YOLOv12, and enhanced HOA presents better performance, which validates that all modules contribute to a significant pathway. Thus, the results recommend that integrating cross-attention based fusion with robust recognition and temporal learning improves performance in multi-sensor ADAS scenarios, particularly under complex driving environments.

6. Conclusion and Future Work

In this manuscript, the CAMSF-ADAS approach has been presented, which incorporates multimodal sensor data attained from camera, LiDAR, and radar for effective perception and object recognition. By calibrating the camera, LiDAR, and radar together in space and time, all sensor data is aligned perfectly for better accuracy and synchronization. In addition, an MHCAFm is introduced for proficiently integrating complementary information from different sensing modalities. The YOLOv12 model mapped and classified adjacent road objects in diverse driving conditions. Additionally, a TFT classifier is utilized to model sequential dependencies and enhance classification accuracy. The hyperparameters of the TFT model are refined using the IHOA, enhancing both classification accuracy and convergence efficiency. Empirical validation indicated that the proposed CAMSF-ADAS system yields superior results than existing methods. Consequently, it serves as a highly capable solution for resilient multi-sensor object detection within ADAS systems. The present study mostly concentrates on strong object detection and categorization utilizing LiDAR, radar, and camera information in ADAS surroundings. Although the proposed method achieved a promising outcome, further developments can be investigated. In the future, the technique can be applied in real-world vehicle platforms to measure its practical applicability under dynamic driving situations. Additionally, the approach can be encompassed with additional sensor modalities and larger real-time datasets for enhancing generalization ability under different environmental settings. Future research may also examine vehicle-to-infrastructure communication and online adaptive learning for improving perception outcomes in intelligent transportation technologies.

Declarations

Ethics approval and consent to participate: Not applicable.

Consent for publication: Not applicable.

Availability of data and materials:

The data that support the findings of this study are openly available in Kaggle repository at <https://www.kaggle.com/datasets/mitanshuchakrawarty/nuscenes> and <https://arthurouaknine.github.io/codeanddata/carrada>, reference number [27, 28].

Competing interests:

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding: Not applicable.

Clinical trial number: Not applicable.

Authors' contributions:

RM: Conceptualization, Methodology, Investigation, Software, Validation, Writing - Original Draft.

PKR: Supervision, Writing - Review & Editing.

BS: Data curation, Resources.

GKP: Data curation, Resources.

CK: Data curation, Resources.

TDR: Data curation, Resources.

CSA: Data curation, Resources.

Acknowledgements: Not applicable.

References

1. William, P., Darwante, N.K., Pawar, A.B., Jawale, M.A., Verma, A.: Framework for Implementation of Smart Driver Assistance System Using Augmented Reality. In: Marriwala, N., Tripathi, C.C., Jain, S., and Kumar, D. (eds.) *Mobile Radio Communications and 5G Networks*. pp. 365–383. Springer Nature Singapore, Singapore (2023)
2. Mahmoud, U.S., Albahri, A.S., AlSattar, H.A., Zaidan, A.A., Talal, M., Mohammed, R.T., Albahri, O.S., Zaidan, B.B., Alamoodi, A.H., Qahtan, S.: DAS benchmarking methodology based on FWZIC II and FDOSM II to support industrial community characteristics in the design and implementation of advanced driver assistance systems in vehicles. *J Ambient Intell Human Comput.* 14, 12747–12774 (2023). <https://doi.org/10.1007/s12652-022-04201-4>
3. Nadeem, H., Javed, K., Nadeem, Z., Khan, M.J., Rubab, S., Yon, D.K., Naqvi, R.A.: Road Feature Detection for Advance Driver Assistance System Using Deep Learning. *Sensors.* 23, 4466 (2023). <https://doi.org/10.3390/s23094466>
4. Rana, K., Khatri, N.: Automotive intelligence: Unleashing the potential of AI beyond advance driver assisting system, a comprehensive review. *Computers and Electrical Engineering.* 117, 109237 (2024). <https://doi.org/10.1016/j.compeleceng.2024.109237>
5. Kasarapu, S., Dinakarrao, S.M.P.: Performance and Environment-Aware Advanced Driving Assistance Systems. *IEEE Trans. Comput.* 74, 131–142 (2025). <https://doi.org/10.1109/TC.2024.3475572>
6. Ali, W.A.-E., Saleh, H.A., Eid, H.S., Fathi, A.M., Mohamed, A.-E.R.A.: Design and Implementation of Advanced Driver Assistance Systems. In: 2024 International Telecommunications Conference (ITC-Egypt). pp. 1–6. IEEE, Cairo, Egypt (2024)
7. Adhikary, S., Seethamasetti, A., Rajkumar, S.: AWS Integrated Machine Learning Architecture for Advanced Driver Assist System (ADAS). In: 2025 International Conference on Sensors and Related Networks (SENNET) Special Focus on Digital Healthcare(64220). pp. 1–6. IEEE, Vellore, India (2025)
8. Utane, A., Mohod, S., Thakare, Y., Ismail, F.S.M.: An ensemble learning with deep feature extraction approach for recognition of traffic signs in advanced driving assistance systems. *J Integr Sci Technol.* 13, 1133 (2025). <https://doi.org/10.62110/sciencein.jist.2025.v13.1133>
9. Lee, G., Lee, K., Hou, J.-U.: Classifying Advanced Driver Assistance System (ADAS) Activation from Multimodal Driving Data: A Real-World Study. *Sensors.* 25, 6139 (2025). <https://doi.org/10.3390/s25196139>
10. Razeena, C.P., Simon, P.: Deep Learning based Automated Braking Decision-making for Advanced Driver Assistance System. *Procedia Computer Science.* 258, 552–562 (2025). <https://doi.org/10.1016/j.procs.2025.04.290>
11. Shi, R., Zhang, S., Wu, Y.: Research on Intelligent Evaluation Method of High Order Advanced Driving Assistance System Based on Target Object Information. In: *Proceedings of the 2024 International Conference on Intelligent Driving and Smart Transportation*. pp. 171–176. ACM, Guangzhou China (2024)
12. Yang, E., Yi, O.: Enhancing Road Safety: Deep Learning-Based Intelligent Driver Drowsiness Detection for Advanced Driver-Assistance Systems. *Electronics.* 13, 708 (2024). <https://doi.org/10.3390/electronics13040708>
13. Falaschetti, L., Manoni, L., Palma, L., Pierleoni, P., Turchetti, C.: Embedded Real-Time Vehicle and Pedestrian Detection Using a Compressed Tiny YOLO v3 Architecture. *IEEE Trans. Intell. Transport. Syst.* 25, 19399–19414 (2024). <https://doi.org/10.1109/TITS.2024.3447453>
14. Gomes, D.T., Tamariz, A.D.R.: A Driver Assistance System Based on YOLO Object Detection: Development and Experimental Validation in the CARLA Simulator. *JBCS.* 32, 1250–1257 (2026). <https://doi.org/10.5753/jbcs.2026.6503>
15. Murthy, J.S., Siddesh, G.M., Lai, W.-C., Parameshachari, B.D., Patil, S.N., Hemalatha, K.L.: ObjectDetect: A Real-Time Object Detection Framework for Advanced Driver Assistant Systems Using YOLOv5. *Wireless Communications and Mobile Computing.* 2022, 9444360 (2022). <https://doi.org/10.1155/2022/9444360>
16. Hong, S.M., Park, J.-W., Jeong, J.-H., Kim, S.Y.: Real-Time Defect Detection of Capacitive Touch Pads for Hands-Off Detection in Advanced Driver Assistance Systems. *Applied Sciences.* 16, 1675 (2026). <https://doi.org/10.3390/app16041675>
17. Sivakumar, S., Tripathi, S.: Real time object detection for advanced driver assistance systems using deep learning techniques. *IJECE.* 15, 4942 (2025). <https://doi.org/10.11591/ijece.v15i5.pp4942-4953>
18. Amin, R.A., Hasan, M., Wiese, V., Obermaisser, R.: FPGA-Based Real-Time Object Detection and Classification System Using YOLO for Edge Computing. *IEEE Access.* 12, 73268–73278 (2024). <https://doi.org/10.1109/ACCESS.2024.3404623>

19. Shahzad, M.Z., Hanif, M.A., Shafique, M.: Robust ADAS: Enhancing Robustness Of Machine Learning-Based Advanced Driver Assistance Systems For Adverse Weather. In: 2024 IEEE International Conference on Image Processing Challenges and Workshops (ICIPCW). pp. 4240–4246. IEEE, Abu Dhabi, United Arab Emirates (2024)
20. Thakker, Z.L., H. Buch, S.: Effect of Feature Scaling Pre-processing Techniques on Machine Learning Algorithms to Predict Particulate Matter Concentration for Gandhinagar, Gujarat, India. IJSRST. 410–419 (2024). <https://doi.org/10.32628/IJSRST52411150>
21. Lyu, X., Liu, S., Qiao, R., Jiang, S., Wang, Y.: Camera, LiDAR, and IMU Spatiotemporal Calibration: Methodological Review and Research Perspectives. *Sensors*. 25, 5409 (2025). <https://doi.org/10.3390/s25175409>
22. Zhao, L., Ke, C., Jia, Y., Xu, C., Teng, Z.: Infrared and Visible Image Fusion via Residual Interactive Transformer and Cross-Attention Fusion. *Sensors*. 25, 4307 (2025). <https://doi.org/10.3390/s25144307>
23. Ge, T., Ning, B., Xie, Y.: YOLO-AFR: An Improved YOLOv12-Based Model for Accurate and Real-Time Dangerous Driving Behavior Detection. *Applied Sciences*. 15, 6090 (2025). <https://doi.org/10.3390/app15116090>
24. Song, J., Zhu, S., Li, W., Xue, G., Gao, X.: A novel robust heating load prediction algorithm based on hybrid residual network and temporal fusion transformer model. *Energy*. 318, 134779 (2025). <https://doi.org/10.1016/j.energy.2025.134779>
25. Taha, S.A., Abdulsada, M.A., Mohamed, M.A.E., Alruwaili, M., Emara, A.: Enhanced maximum power point tracking using hippopotamus optimization algorithm for grid-connected photovoltaic system. *Sci Rep*. 16, 9991 (2026). <https://doi.org/10.1038/s41598-026-40918-4>
26. Pei, S., Sun, G., Tong, L.: An improved hippopotamus optimization algorithm based on adaptive development and solution diversity enhancement. *PeerJ Computer Science*. 11, e2901 (2025). <https://doi.org/10.7717/peerj-cs.2901>
27. chakrawarty, mitanshu: Nuscenese Dataset, <https://www.kaggle.com/datasets/mitanshuchakrawarty/nuscenes>
28. Ouaknine, A., Newson, A., Rebut, J., Tupin, F., Perez, P.: CARRADA Dataset: Camera and Automotive Radar with Range- Angle- Doppler Annotations. In: 2020 25th International Conference on Pattern Recognition (ICPR). pp. 5068–5075. IEEE, Milan, Italy (2021)
29. Ahmad, J., Del Bue, A.: mmFUSION: Multimodal Fusion for 3D Objects Detection, <https://arxiv.org/abs/2311.04058>, (2023)
30. Brödermann, T., Sakaridis, C., Dai, D., Van Gool, L.: HRFuser: A Multi-Resolution Sensor Fusion Architecture for 2D Object Detection. In: 2023 IEEE 26th International Conference on Intelligent Transportation Systems (ITSC). pp. 4159–4166. IEEE, Bilbao, Spain (2023)
31. Liang, M., Yang, B., Wang, S., Urtasun, R.: Deep continuous fusion for multi-sensor 3D object detection. In: Proceedings of the european conference on computer vision (ECCV) (2018)
32. Liang, M., Yang, B., Chen, Y., Hu, R., Urtasun, R.: Multi-task multi-sensor fusion for 3D object detection. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR) (2019)