Design and Analysis of a Hybrid Spiking–Deep Neural Network for Energy-Efficient Object Detection

by

Asrar Azim
21301634
Md. Arif Uz Zaman Sakib
21301318
Md. Rafid Jamal
22341023
Nabiha Binte Yousuf
21201752
Shihana Sultana Jhilik
21301403

A thesis submitted to the Department of Computer Science and Engineering
in partial fulfillment of the requirements for the degree of
B.Sc. in Computer Science and Engineering

Department of Computer Science and Engineering
Brac University
January 2026.

Declaration

It is hereby declared that

1. The thesis submitted is my/our own original work while completing degree at Brac University.

2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.

3. The thesis does not contain material which has been accepted, or submitted, for any other degree or diploma at a university or other institution.

4. We have acknowledged all main sources of help.

Student's Full Name & Signature:

Asrar

Asrar Azim
21301634

Arif Uz Zaman

Md. Arif Uz
Zaman Sakib
21301318

Rafid

Md. Rafid
Jamal
22341023

Nabiha

Nabiha Binte
Yousuf
21201752

Jhilik

Shihana
Sultana Jhilik
21301403

Approval

The thesis/project titled "Design and Analysis of a Hybrid Spiking–Deep Neural Network for Energy-Efficient Object Detection" submitted by

1. Asrar Azim (21301634)
2. Md. Arif Uz Zaman Sakib (21301318)
3. Md. Rafid Jamal (22341023)
4. Nabiha Binte Yousuf (21201752)
5. Shihana Sultana Jhilik (21301403)

of Spring, 2025 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of B.Sc. in Computer Science in 2026 Year.

Examining Committee:

Supervisor:
(Member)

Mr. Rafeed Rahman

Senior Lecturer
Department of Computer Science and Engineering
Brac University

Co-Supervisor:
(Member)

Dr. Md. Golam Rabiul Alam

Professor
Department of Computer Science and Engineering
Brac University

Thesis Coordinator:
(Member)

____

Dr. Md. Golam Rabiul Alam

Professor
Department of Computer Science and Engineering
Brac University

Head of Department:
(Chair)

____

Sadia Hamid Kazi

Chairperson and Associate Professor
Department of Computer Science and Engineering
Brac University

Abstract

In autonomous vehicular systems, object detection performs a fundamental task in which real-time processing and energy efficiency are crucial for execution on edge devices. Faster R-CNN, which are conventional deep neural network based detectors, provide high accuracy but at the same time sustain substantial power consumption because of continuous and computationally intensive operations. This research proposes the design of a hybrid spiking deep neural network that is energy efficient in object detection by integrating the architecture of SNN (Spiking Neural Network) that uses LIF (Leaky Integrate and Fire) neurons and temporal spike processing into the feature extraction stage, reducing redundant neural activities simultaneously preserving effective object detection functionality by utilizing event-driven spiking computation. The framework is assessed on the KITTI dataset which comprises real world scenarios of autonomous driving. Experimental evaluation shows that it achieves a mAP@0.5 of 0.7395, a mAP@0.7 of 0.5926 and a COCO style mAP@[0.50:0.95] of 0.4771 along with high recall and robust localization. Moreover, the hybrid model enhances detection accuracy over regular SNN models and improves recognition of small and distant objects. These results validate that the presented model has a potential and promising aspect for constructing energy efficient object detection systems for autonomous and edge based deployment.

Keywords: Spiking Neural Network (SNN), Faster R-CNN, Object detection, Energy efficiency , KITTI dataset, edge computing, Autonomous driving.

Table of Contents

List of Figures

Chapter 1

Introduction

1.1 Background

Autonomous vehicle is a type of technology, where a car can make its own decision while driving. Such vehicles are capable of responding effectively to traffic hazards, whether by stopping for pedestrians or obstacles or by making real-time decisions to avoid potential collisions. The integration of such vehicles aims to enhance modern lifestyles by reducing the burden of decision-making during driving. For instance, momentary lapses in driver concentration can result in serious accidents, a risk that autonomous systems are designed to minimize. Autonomous vehicles are something that can help to guess what will happen in future and notify the driver beforehand or it can drive it by itself [1]. Additionally, there will be less of a traffic hazard because autonomous vehicles are completely programmed where they know how to make the best decisions while driving. It is better for new drivers, so it will make their life a lot easier and will improve the traffic system. Moreover, in situations where people who are not supposed to drive cars such as someone who doesn't have their driving lessons yet, or someone with disabilities who cannot function a car, those people are someone who can be highly beneficial from this modern technology.

However, autonomous vehicles are still yet to develop completely. Currently, few tech companies have run some tests, where the accuracy of the autonomous vehicles weren't satisfying as different people were injured. Additionally, it raises a question like, if there were a situation where the car has to save either one person or a group of people, which decision will it take? Moral questions like that need to be answered before it is available for the consumers. Moreover, there is concern whether it will be energy efficient and environmentally friendly or not? Autonomous vehicles are fully electric and use highly powerful sensors and cameras to detect objects. The fact that they do not consume fuel does not mean that they are energy-efficient. Still, when powered by renewable energy sources, autonomous cars can be more effective and eco-friendly than conventional cars.

1.2 Rational of the study or motivation

In computer vision, object detection mainly deals with locating and spotting objects by viewing images or videos. Thanks to their ability to find complex features from images without any training, Convolutional Neural Networks especially have changed the field. They focus on achieving reliable results in difficult situations, more than the traditional approaches that feature are engineered beforehand. Checking on surveillance projects, yielding fine results in field like self-driving vehicles, medical imaging, and business analytics, these DNN models like YOLO, SSD, and Faster R-CNN are highly effective. They are skilled at working with different sizes, light conditions, and how objects are positioned, making it simple to analyze large sets of information with powerful GPUs [2].

It is challenging to deploy DNNs in energy-limited regions since they take a lot of computing power. For this reason, researchers are now looking into Spiking Neural Networks (SNNs). SNNs reproduce the type of communication that happens in the brain by quickly sending out bursts of information labeled as spikes rather than slower and regular data. Being event-driven, this type of computing uses less electric power and complements hardware created for performing low-power operations. Traditionally, SNNs were unable to complete complex tasks such as object detection as well as DNNs because of problems with training and spike computation. Now, improvements and advancements in this field are closing the gap between the two. The major goal of SNNs for object detection is to reach similar accuracy as DNNs, but while using far less energy, which allows for quick detection of objects in mobile phones and other devices.

1.3 Problem statement

Autonomous vehicles are self-driving cars powered by artificial intelligence and this makes them move safely to their destinations. These vehicles take measurements in the form of cameras, LiDAR, radar and other sensors to know what is around them. This information assists the car to recognize the objects, comprehend the traffic conditions and prevent accidents. Machine learning models then process the collected data to assist the vehicle in making decisions in driving.

Deep Neural Networks (DNNs) are widely applied in self-driving cars to perform such tasks as object recognition, lane recognition, and predicting traffic. DNNs are based on the ability to process images and other sensor input continuously. They examine a number of frames per second with multiple layers of artificial neurons. Even though DNNs are very accurate, they consume a lot of power since they are constantly processing data even when the environment is not changing. This consumes a lot of energy, which makes DNNs inappropriate to vehicles with small computing and power capabilities. However, Spiking Neural Networks (SNNs) do not operate like DNNs. They are driven by the way the human brain processes information. SNNs are not processing data continuously but instead, they just respond to a change in input. They transmit information by means of spikes and as such, they can only process information when something of importance occurs like movement on the scene. Due to this event-based behavior, the SNNs use much less energy and are most effective when these are applicable in real time.

However, the basic designs of SNN are not as efficient as DNNs for complex vision tasks. To overcome this, hybrid systems that make use of SNNs and DNNs are utilized whereby the power efficiency of SNNs is used together with the good precision of deep networks. This paper incorporates a spiking neural network into a Faster R-CNN system to detect the objects. The main problem lies in the question of how to develop a fast and energy efficient hybrid model capable of being able to handle visual or event based data on resource limited devices like those utilized in autonomous systems. The proposed approach will ensure that it is supported with a high level of accuracy and consumes less power by integrating the strengths of SNNs and DNNs, which makes it applicable to real-world edge applications.

1.4 Objective

The primary task of the project is to design, program and test a fast and energy-efficient hybrid Spiking-Deep Neural Network (SNN-DNN) object detector. Object detection is a very important aspect of computer vision and is commonly applied in autonomous vehicles, robotics, surveillance systems, and other intelligent systems. These systems need models that have the capability of making rapid and correct decisions as well as use minimum power.

Object detection in traditional Deep Neural Networks (DNNs) has been performed well since it is able to learn rich features on large datasets. Nevertheless, DNNs are constantly running computations and need an extensive amount of computing resources, which is why it makes them less applicable to devices with constraints. Spiking Neural Networks (SNNs) are event-based processors of information, thereby being very energy efficient. Nevertheless, SNNs themselves can be affected by the challenges including reduced accuracy on complex vision tasks [3].

To overcome these shortcomings, the proposed study is aimed at incorporating SNN into a Faster R-CNN architecture to combine the low power consumption of SNNs with the high detection rates of DNNs. The purpose of the project is to enhance speed, precision, and energy efficiency during the object detection tasks, which suggests the model is applicable when used on edge and low-power devices.

This research aims to find out:

- To develop a hybrid SNN-DNN Faster R-CNN model that is applicable to real-time detection of objects.
- To detect at high accuracy and still increase the speed of inference.
- To test the presented hybrid model with conventional object detection dataset.
- To examine the model performance as measured by accuracy, latency, energy consumption, and model size.
- To show that hybrid SNN-DNN models are suitable in the use of low-power and edge devices.

The study will continue to make a contribution towards the creation of viable, quick and energy saving vision systems that can be implemented in practice in autonomous and embedded implementations.

1.5 Methodology in brief

The methodology of this study can be split into four main steps which are dataset preparation, Model implementation, training and evaluation and comparative analysis. The figure 1.1 illustrates our work flow in sequence.

1. Dataset Preparation
The benchmark KITTI dataset is utilized in this work, and it is dedicated to autonomous driving and includes annotated real-world street scene data. The dataset includes objects like pedestrians, cars, bicycles etc. In addition, to suit our research purpose, we have prepared the dataset to both training and testing splits. In addition, there are scaling, flipping, normalization etc data augmentation strategies to augment power of generalization.

2. Model Implementation
This paper compares two streams of object detection, a regular DNN-based Faster R-CNN and a hybrid model of Faster R-CNN with SNN. The use of this design will incorporate the energy performance of SNNs and the performance of DNNs when it comes to detecting.

3. Training and Evaluation
The two models are also trained and tested on the identical split of the data to enable them to be compared fairly. It is evaluated in the parameters of the normal object detection metrics, the mean Average Precision (mAP), Intersection over Union (IoU), and the inference latency. The backpropagation of the DNN elements and the spike-based processing rules determine the training of the SNN backbone respectively.

4. Comparative Analysis
The two models are contrasted to each other to evaluate the detection accuracy, inference speed, energy usage and complexity of the model. As this discussion will determine, this is the performance-power consumption trade-off followed by the suitability of the hybrid SNN-DNN model to execute on edge and resource-constrained systems (autonomous and real-time systems).

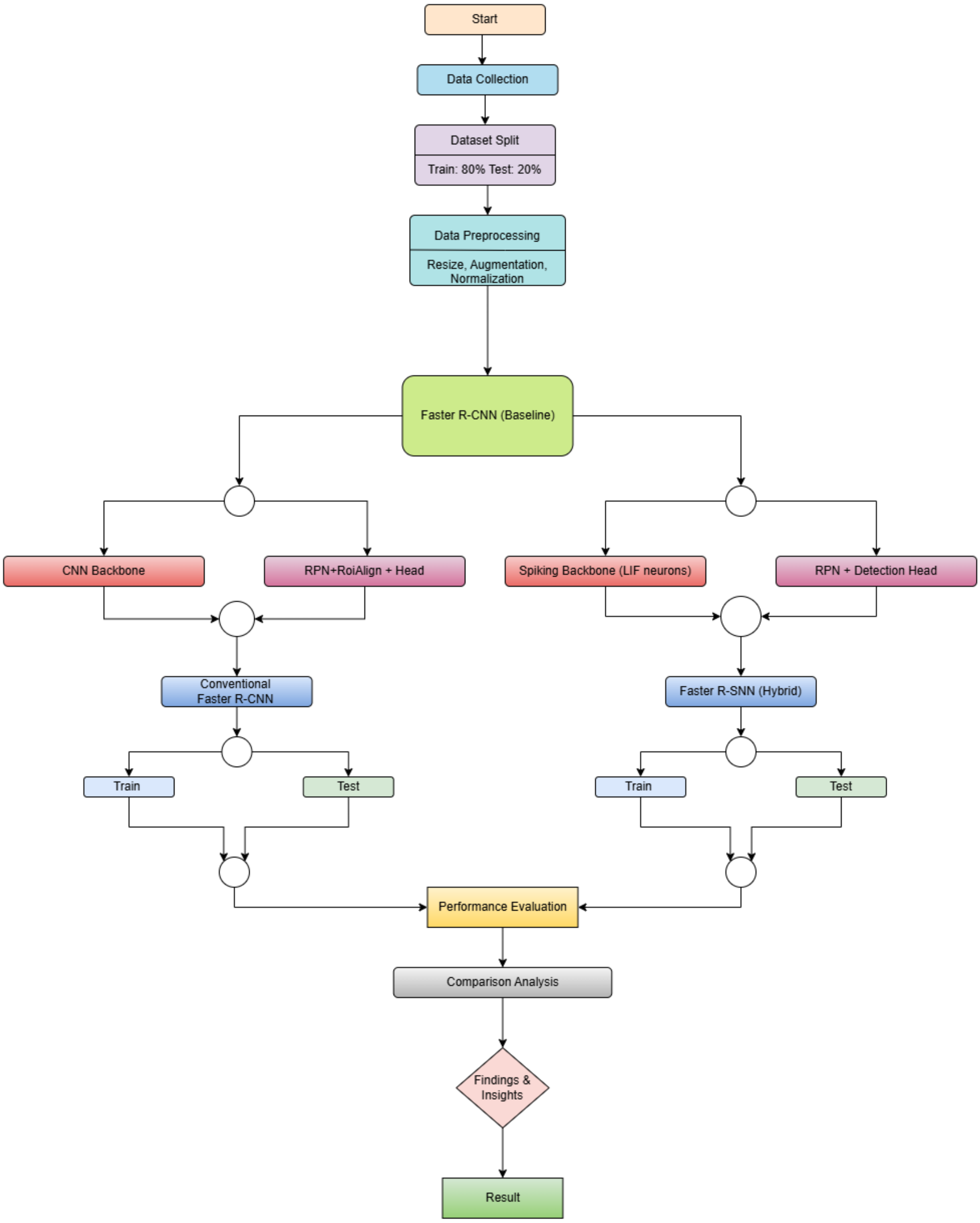

Figure 1.1: Work Plan

1.6 Research contribution

The proposed thesis has a contribution to the field of energy-efficient computer vision. Rather than completely superseding the conventional deep learning models, this work replaces the chosen internal units of the Faster R-CNN architecture and incorporates the spiking neural concepts. It is an efficient way to minimize energy consumption and still achieve reasonable object recognition.

The primary findings of this study are as follows:

- Creation of spiking computation hybrid Faster R-CNN: This paper demonstrates that it is possible to convert Faster R-CNN to a hybrid SNN-based model without altering its general detection framework. The key steps like feature extraction, region proposal generation and object classification are kept, and internal layers are transformed to spike-based processing.

- Object detection design that is energy efficient: The given model makes use of the spiking neurons which process information on the basis of discrete events, rather than the continuous values. Such behavior of events can minimize the amount of computing that should be done and reduce power usage, making the model an appropriate model for edge devices and low-power vision.

- Better accuracy in comparison with fully spiking models: Compared to a fully spiking neural network, the hybrid design is simpler to detect objects more accurately, since it uses both traditional convolutional layers and spiking computation. This trade-off enables the model to have the advantage of spiking based performance without the loss of performance frequently observed in entirely SNN-based detectors.

- Better detection of small and remote objects: combination.Height-based discussion reveals that the Hybrid R-SNN maintains more detections with small or distant objects than the traditional Faster R-CNN model does. This is an advantage that exists due to the temporal characteristics of spiking neurons which enables weak visual signals to build up over time to be detectable. This conclusion demonstrates the appropriateness of hybrid spiking models to challenging cases of detection.

- Accuracy/energy Trade-off analysis: The thesis gives an analysis of the trade-off between performance in detection and energy-saving. The findings indicate that good detection accuracy may be obtained at lower computational cost, and this provides the practical importance of hybrid models.

- Enhanced memory of safety-critical applications: The hybrid model has better recall, as shown by object classes, that is, it will not be more likely to fail to see objects in a scene. This property is significant in systems where safety is a crucial concern like autonomous vehicles and smart surveillance, where missing a target can be dangerous.

- Generalization of SNNs to complete object detectors: This study broadens the application of spiking neural networks of simple classification to a full system of object detection. The suggested hybrid solution paves way to new research opportunities on developing scalable and energy-conscious vision models on resource-constrained systems.

1.7 Scopes and challenges

The design and testing of a hybrid Spiking-Deep Neural Network (SNN-DNN) that would assist in the detection of objects in real-time and in the autonomous systems is of concern in the proposed study. The model that will be generated by the researchers through the research they conduct will as well assist in creating a model that is accurate, fast and also consumes less energy and can be applicable on edge computing machines with a low level of computing capabilities.

The hybrid model incorporates an SNN backbone in the architecture of Faster R-CNN, where the event-driven behavior of spiking neurons makes use of the idle computation. Performance evaluation is performed on the basis of the accuracy, inference speed, energy consumption, and model size with regard to the standard object detection data. It also examines how the model can be applied in dynamic real-world systems e.g. autonomous vehicles where efficient and reliable object detection is critical.

Challenges of Developing a hybrid SNN-DNN system of object detection:

- Training complexity- SNNs use spike information, therefore, it makes optimization and learning even more challenging than general DNNs.
- Performance versus efficiency tradeoff - To achieve high accuracy and lower latency and power consumption this has to be designed carefully.
- Lack of good software and hardware support - SNNs tools are younger than DNN frameworks and neuromorphic hardware is not commercially available.

To address these concerns, one needs a system of object detection developed with practical, fast, and energy efficiency to be employed in edge and autonomous applications.

Chapter 2

Literature Review

2.1 Preliminaries

Particularly with autonomous driving and advanced driver assistance systems (ADAS), the integration of artificial intelligence in vehicle systems has become a major focus of research. Object detection is the core focus in autonomous vehicles. The choice of neural network architecture becomes crucial as vehicular systems significantly rely on edge computing architectures to meet real-time processing requirements while reducing latency and maintaining efficiency. Thus we decided to examine the comparative landscape between Spiking Neural Networks (SNNs) and Deep Neural Networks (DNNs) for object detection applications specifically focused for edge devices in vehicular environments. Spiking Neural Networks represent an example toward biologically-inspired computing, processing information through discrete spikes rather than continuous values. This approach offers theoretical advantages in energy efficiency, making SNNs particularly attractive for battery-powered edge devices in vehicular applications. But the application of SNNs to object detection remains in relatively early stages compared to DNNs. Recent research has explored the conversion of pre-trained DNNs to SNNs, as well as direct training of SNNs for computer vision tasks. On the other hand, deep Neural Networks have dominated the object detection landscape in vehicular applications, with architectures such as YOLO, SSD, and R-CNN variants being extensively adapted for automotive use cases. Research has demonstrated that DNNs can achieve high accuracy in detecting vehicles, pedestrians, traffic signs, and road markings under various environmental conditions. While DNNs have established dominance in vehicular object detection, the theoretical advantages of SNNs in terms of energy efficiency and event-driven processing present compelling opportunities for edge deployment. Thus this research aims to address these aspects by providing a thorough comparative study that evaluates both models under realistic vehicular edge computing constraints. Therefore, in this section, we also review some of the past research papers where all these techniques are used and analyze their results, limitations, and improvements in the future.

2.2 Review of existing research

ECSLIF-YOLO combines AI frameworks of ECS-LIF spiking neurons and the YOLO to enhance object recognition in autonomous driving [4]. Extracellular space dynamics leads to a more stable, robust, and biological SNNs. Performances on BDD100K and KITTI reached a peak of 0.917 at real-time (¿75 FPS) performance, which performs better than previous SNN-based models under adverse conditions. Nevertheless, the model operates based on theoretical predictions of energy and presents a computational complexity, which prevents its practical implementation unless optimized further in its hardware.

In [5], the Time-of-Flight (ToF) sensor system was also suggested as a distributed solution to LiDAR in autonomous driving. The system has seven synchronized ToF sensors, which cover 360deg, and fused point clouds are processed with an SNN via temporal voxel coding to achieve 65% mAP on real-world data. The system is able to identify nearby objects (0.5-6 m) and produce occupancy grids with reduced latency and better-blind spot awareness. Nevertheless, it does not work in high sunshine, the sensor range is low in comparison with LiDAR, and multi-sensor registration drift and data shortage is one of the major issues.

Study [6] suggested SFDNet, a complete spiking based object detector model which combines RGB images and event cameras to create power-saving car vision. It presents a neuron model, Leaky Integrate-and-Multi-Fire, and a two-pathway network with a spiking YOLOX detection head to process the entire system. Assessments on both PKU-DAVIS-SOD and DSEC-Detection datasets demonstrated state-of-the-art results with low-light and motion-blur conditions and using up to 32x less energy than all-conventional solutions. Nevertheless, the task of detection of small or occluded objects is very difficult as it has a class imbalance and may also lose information during the event-to-image conversion.

The robustness of automotive object detectors under camera artifacts and video compression artifacts was studied in Study [7]. The application of KITTI MoSeg with Faster R-CNN and YOLOv5 demonstrated that compression alone did not have significant effects but could achieve up to 90 per cent higher accuracy in the absence of serious noise and hindrances. The augmented data retraining enhanced the robustness by up to 35 percent particularly during moderate degradation. Nevertheless, an extreme occlusion (¿80%), and none of the real-world degraded datasets address evaluation realism.

In [8], RT-SNN was put forward as a real-time scheduling scheme of SNN-based object detection in self-driving cars. The system makes the timesteps dynamic and recycles membrane potentials using a membrane confidence measure, thus enhancing speed and energy consumption. Real-time performance Spiking-YOLO experiments on KITTI showed to be up to 280x or more energy-efficient than ANN models. However, the method is restricted to the SNN structures and based on the rigid sequence of tasks, which can decrease plasticity in a complicated situation.

The first SNN to be developed on 4D radar-based 3D object detection was SpikingRTNH proposed in Study [9]. It proposes Biological Top-Down Inference (BTI) to decrease noise and improve the detection accuracy in addition to increasing energy efficiency. When tested on the K-Radar data in a variety of weather conditions, the model gave 51.1% AP3D, 57% APBEV, and lower energy use by 78%. Nevertheless, it is based on the implementation of GPU and demonstrates average performance deterioration as compared to the ANN baselines.

Saeedizadeh et al. offer a comprehensive overview of existing deep learning-based 2D object detection techniques in AD, specifically in vehicle and pedestrian detection [10]. In their study, they discuss and classify existing object detection methods into one-stage and two-stage object detection, including YOLO, SSD, and R-CNN, and their performance in adverse situations such as occlusion and illumination variations. More than 90 object detection models and 18 publicly available datasets, including KITTI and nuScenes, are considered in their study.

A hybrid SNN-ANN model that detects events of an object in an automotive setting with an attention-based bridge module was proposed in study [11]. This module transforms the sparse spike information into dense features maps using spatial-temporal attention, which allows the ANN processing. The Gen1 and Gen4 experiments recorded competitive mAP of 0.35 and 0.27 using a small number of parameters 6.6M and are highly energy and parameter efficient. Even though performance can be validated real time on Intel Loihi 2, the attention module introduces computational overhead and integration costs, and thus needs additional optimization to implement neuromorphic systems entirely.

Current SNN detectors such as Spiking-YOLO are time consuming, which restricts real-time edge deployment. To make timestep compression and STDI coding under timestep compression more efficient in energy-efficient SNN, Study [12] suggested SUHD. Better accuracy with 2000x/750x fewer timesteps on PASCAL VOC and 300x/150x fewer timesteps on MS COCO, and as many as 400x lower energy usage was achieved with the converted YOLOv5s model. Nevertheless, spike conversion can be incapable of preserving temporal information and it does not fully generalize to all neuron models e.g. LIF.

The authors presented a Spiking Feature Pyramid Network (SpikeFPN) for efficient event based object detection using less computing power [13]. They proposed a spike-triggered adaptive threshold and surrogate gradient training to ensure stable learning while examining membrane potential dynamics. By using the GEN1 Automotive Detection dataset, the model gained 0.477 mAP which demonstrates a 9.7% improvement over previous SNN baselines and exceeding advanced ANNs and traditional SNNs. The model heavily relies on sensor data which constraints its generalization and future expandability in spite of its efficient and robust design.

This study emphasizes event based cameras for vehicular object detection, highlighting their low latency and low power consumption in comparison with conventional cameras [1]. By utilizing datasets such as GEN1, DDD17, DSEC, PKU-DDD17 and related benchmarks, It contrasts DNN, SNN, GNN and multimodal methods. This survey demonstrates that while SNNs are energy efficient and effective in classification, they still fall behind DNNs in regression based object detection tasks. It also notes challenges of event data which comprises irregular timing and poor performance for static objects, indicating sensor fusion with conventional cameras as a solution.

The authors presented SPLEAT, a neuromorphic accelerator developed to efficiently implement spiking neural networks on low power hardware [14]. Each spiking layer has a dedicated neural processing unit activated only by spikes, facilitating batch norm fusion and multiple neuron models such as IF, LIF and PLIF. The model is evaluated on the Prophesee GEN1 dataset utilizing a 32-ST-VGG network which gained effective object detection with 1.08M parameters using only 490 mJ per prediction. SPLEAT executes 43% faster with a microcontroller, it's in 3.1 MB memory, consumes 1 W power and finishes tasks under 1 s, even though implementation losses still need improvement.

In [15], the authors propose an energy-efficient FPGA accelerator for SNN-based object detection models like Spiking-YOLO, which are too large to be deployed on FPGAs in real-time. Algorithmic optimizations include channel pruning, batch normalization fusion, and scale-aware quantization. In addition, a fully pipelined hardware architecture is proposed. Systolic arrays and inference neurons are also used to improve performance. Experimental results demonstrate a 26x model size reduction while retaining accuracy and achieving real-time performance at 681 fps with low power consumption.

Laboni and Abichandani conducted a survey of 151 studies on Event-based Spiking Neural Networks (SNNs) in the context of object detection in computer vision [16]. The survey paper presents the advantages of using Event-based Spiking Neural Networks in terms of energy and low latency. The survey paper covers Event-based and RGB data sources, spike coding schemes, SNNs, and learning algorithms like STDP and ANN-to-SNN conversion.

In the research paper titled "Autonomous Driving with Spiking Neural Networks", the first unified Spiking Neural Network (SNN) framework for end-to-end autonomous driving with strict energy constraints, is proposed [17]. SAD is a Spiking Neural Network-based model for autonomous driving, consisting of spiking neurons for perception, prediction, and planning. SAD creates spatiotemporal bird's-eye-view representations, predicts the future using dual-pathway temporal mixing, and makes driving plans using spiking recurrence. SAD is competitive with ANN-based methods on the nuScenes dataset while providing lower energy consumption, making SNN-based methods suitable for autonomous driving.

R-SNN was presented in [18] as a region-based SNN to detect 3D objects that are 3D hardware-optimized. The model transforms an R-CNN (VGG-16) into a spiking model by weight normalization with the help of IF and mirror neurons with linear regression decoding. It was tested on VOC 2007 and scored 63.1% mAP and surpassed Spiking-YOLO by using only 6 W on Darwin Mouse hardware. However, it is limited in terms of its performance because of rate-coding noise and the reliance on the quality of the original DNN when converting it.

This research proposes a framework integrating spiking neural networks (SNNs) with event based cameras for detecting objects efficiently [19]. It utilizes event encoding, Surface of Active Events (SAE), LIF based SpikingYOLOv4 and semi automatic labeling to extend training data without full manual annotation. The method facilitates real-time, low power detection in challenging environments and advantages from CNN to SNN conversion for deep SNN modeling. Outcomes validate its scalability and efficiency with future work intended at handling complex datasets and deeper SNN architectures.

This research proposes a fully spiking neural network (SNN) object detector for autonomous driving which is based on Faster R-CNN and trained via surrogate gradients [20]. It attains competitive accuracy on Cityscapes, IDD and BDD datasets at the same time minimizing energy consumption by up to 85% and demonstrating robustness to noisy outputs. The authors present a lightweight approach to detect novel objects utilizing RPN analysis, although it is biased toward known classes. Future work comprises developing fully spiking backbones, enhancing training stability and improving SNN adaptability in dynamic environments.

[21] presented a framework that integrates DNNs with a bicycle motion model for 3D object detection and path production in autonomous trams. It utilizes BiSeNet with Faster R-CNN, FPN based feature extraction and an enhanced Kalman filter which is trained on BDD100K and nuScenes datasets. The system attained high performance (IoU 0.68, mAP on BDD, 0.62 mAP on nuScenes) and precise short term prediction (approximately 1 m error at 1 s), particularly for cars and buses. Nevertheless, performance reduces beyond one second because of limited inter-vehicle interaction modeling and the lack of LiDAR data.

The authors presented a Spiking SiamFC++ framework for object tracking, transforming AlexNet into a spiking neural network and training it utilizing surrogate gradient methods [22]. The model was trained using the GOT-10K dataset and evaluated on OTB2015, UAV123, VOT2016 and VOT2018 benchmarks. It attained strong outcomes which outperformed Siam SNN with 85.24% precision and 64.37% success, matching SiamRPN++ AUC on UAV123 and demonstrating lower precision loss than SiamFC++on OTB. It still falls behind ANN based models in complex scenarios and necessitates further improvement in training complexity and data dependency in spite of superior performance among SNNs.

Spiking Neural Networks (SNNs) are currently considered as an energy-efficient substitute of traditional CNNs to low-power deep learning, especially in autonomous driving [23]. Although CNNs have high accuracy, they are computationally expensive, making them unsuitable to be used in embedded systems, whereas SNNs make use of event-driven processing, sparseness of activation, and asynchronous computation. The developments in the field of training have allowed SNNs to perform more intricate tasks like detection and segmentation of objects. An example of this development is the SpikiLi framework, which transforms CNN-based LiDAR 3D object-detection models into SNNs and is able to run on the KITTI dataset in real-time with competitive accuracy.

In [24], the authors have proposed a Fully Spiking Hybrid Neural Network (FSHN) architecture for efficient object detection in terms of energy consumption. FSHN extends a Spiking RetinaNet with a backbone of ResNet-101 and incorporates both unsupervised STDP training and supervised backprop training. Experimental results on the MS-COCO dataset have shown that FSHN outperforms traditional DNN-based and purely backpropagated spiking neural networks in noisy and low-label scenarios. Replacing MAC operation with accumulation operation results in up to 150 times energy reduction.

Traditional CNN- and LSTM-based autonomous perception systems demonstrate excellent results yet they tend not to be real time synchronized and energy efficient [25]. This has led to a shift of research towards biologically inspired Spiking Neural Networks (SNNs) in terms of event driven and power-efficient properties, but with eventual problems in noisy urban settings. To deal with this, a hybrid Deep spiking neural network based CRF and Probabilistic Particle Filter (PPF) was introduced. The PPF-DSNN model enhances real-time detection and tracking and achieves an accuracy of 96.75 at reduced computational costs which explains its ability to support scalable autonomous vehicle perception systems.

The authors presented a three-stage camera-based deep neural network approach for 3D object detection [26]. The framework utilizes a RPN combining bird's eye view LiDAR features and RGB image features to produce 3D proposals which is followed by a segmentation network and TNet refinement. The framework is evaluated on the KITTI 3D object detection dataset and it attained APs of 69.21% for cars, 43.23% for pedestrians and 55.34% for cyclists. Despite the fusion approach significantly outperforming existing models, further refinement is still necessary for real-world applications.

Faster R-CNN was able to detect objects more quickly than R-CNN and Fast R-CNN, because it eliminated the sluggish external region proposals. The Region Proposal Network (RPN) produces bounding boxes based on shared feature maps and allows end-to-end training and precise and multi-scale localization. The model achieved superior results to selective search based approaches on Pascal VOC and MS COCO with real-time performance comparable to GPUs. Its speed and accuracy balance has enabled it to form a basis of numerous later detection undertakings [27].

2.3 Summary of key findings

Recent advances in spiking neural networks (SNNs) for autonomous driving object detection demonstrate significant potential for energy-efficient perception systems while maintaining competitive performance with traditional artificial neural networks. ECSLIF-YOLO achieved 0.917 mAP on standard datasets with 75+ FPS while drastically reducing energy consumption through biologically-inspired extracellular space dynamics, while hybrid SNN-ANN architectures successfully implemented attention-based bridge modules on Intel's Loihi 2 neuromorphic chip for real-time processing. Novel approaches like SFDNet integrated RGB and event camera data achieving 32× power reduction, and distributed ToF sensor systems provided 360° environmental coverage with 65% mAP for close-range detection. Energy efficiency breakthroughs include RT-SNN's 280× improvement over ANN baselines, SUHD's 750× timestep reduction with 30% accuracy gains, and FSHN's 150× energy savings through accumulated operations. Practical implementations showed promising results with R-SNN achieving 63.1% mAP on neuromorphic hardware consuming only 6W, SPLEAT accelerator enabling first successful low-power chip embedding at 490 millijoules per prediction, and FPGA implementations processing 256×256 images at 681 fps with 4.8W consumption. However, persistent challenges include limited performance under extreme weather conditions, computational complexity for real-world deployment, reduced accuracy for small object detection due to dataset imbalances, and gaps between theoretical energy estimates and actual neuromorphic hardware validation. Despite these limitations, comprehensive surveys analyzing over 151 papers and unified frameworks like SAD demonstrate that SNNs represent a viable path toward sustainable, energy-efficient autonomous driving perception systems, though further optimization for adverse conditions and complete neuromorphic deployment remains necessary.

Chapter 3

Dataset Description

The KITTI dataset [28] is an incredibly popular benchmark dataset that is used to develop an assessment of computer vision algorithms in autonomous driving scenarios. It was surveyed by means of vehicles with several sensors on board like high resolution RGB cameras, LiDAR scanners, GPS, and IMU systems. The dataset includes actual street scenes that were taken in urban, rural, and highway settings; therefore, it is very appropriate in various tasks, including object recognition, object tracking, depth, and understanding of a scene.

The dataset consists of 7,481 images that were captured on the daily sidewalks, streets, and roads. Height of images varies between 370-376 pixels with an average of 374.5 pixels per image. The widths of the images vary between 1224 and 1242 pixels with an average width of 1239.9 pixels per image. The KITTI dataset has 9 classes of objects which are categorised into 9 labels, namely: Pedestrian, Truck, Car, Cyclist, DontCare, Misc, Van, Person_sitting and Tram. All in all, objects are 51,865. The classes are divided in the following way: Car (28,742), DontCare (11,295), Pedestrian (4,487), Van (2,914), Cyclist (1,627), Truck (1,095), Misc (973), Tram (511), and Person_sitting (222). There are a greater number of car cases in the dataset than any other category since cars are the most common items that can be found on highways. There are also fewer instances of other classes such as Tram, Person_sitting since they are not as common when one is on the road [29].

Moreover, the dataset has DontCare regions which depict ambiguous or irrelevant regions in the scene. These areas are not a part of valid object categories and thus do not participate in the loss computation and evaluation as part of a regular KITTI evaluation protocol.

Because of realistic data and exhaustive annotations, the KITTI dataset is commonly utilized when it comes to assessing the strength and real-time capabilities of object detection models. KITTI has been a valuable benchmark in the study of energy-efficient and neuromorphic vision modeling as a way of studying generalization of models to autonomous and edge-based applications in the real world.

Chapter 4

Requirements, Impacts and Constraints

4.1 Final specifications and requirements

This work is dedicated to the design and analysis of a Region based Spiking Neural Network (RSNN)-based framework of object detection to obtain the energy-efficient visual perception. The system demands neuromorphic-inspired modeling with spiking fire neuron dynamics, e.g. Leaky Integrate-and-Fire (LIF) neurons, which process the temporal spike based information. Python is the main programming language, deep learning models based on PyTorch with spiking neural networks (e.g., SpikingJelly or BindsNET) and simulation of spike activity and network behavior are the main software requirements. The model needs benchmark object detection datasets transformed into spike-based representations, and needs to be trained and evaluated with the help of a GPU. Functional requirements comprise correct localization of objects, low-latency inference, and low power use in comparison to deep neural networks used in conventional modes.

4.2 Societal impact

The proposed RSNN-based object detection system has significant societal benefits, particularly in applications where energy efficiency and real-time processing are critical. Low-power intelligent vision systems can be deployed in edge devices such as surveillance cameras, autonomous vehicles, drones, and assistive technologies without relying heavily on cloud-based computation. This enhances accessibility, improves safety, and enables intelligent systems to operate reliably in resource-constrained environments. Moreover, energy-efficient AI solutions contribute to broader societal goals by reducing dependence on high-power computing infrastructure.

4.3 Environmental impact

The conventional deep neural networks demand a lot of computing power, which is energy-consuming and contributes to carbon emission production. The proposed system uses very little energy by taking advantage of RSNNs that can communicate using sparsity in spikes and event-driven processing. This will help in making computing environmentally friendly by reducing the amount of power consumption, as well as increasing the life cycle of hardware products. The use of spiking neural networks contributes to green AI projects and complies with the world-wide policy of reducing the environmental footprint of massive machine learning systems.

4.4 Ethical issues

The main ethical issue associated with this study is the responsible application of AI and the use of data. The use of object detection systems can be used in the situation of surveillance or monitoring and this leads to the issue of privacy and abuse.

The autonomous vehicle systems also provide the possibility of an ethical dilemma when collision is inevitable. As an illustration, when one person is standing alone on one side of the road and a group of people is available on the other side, the car will have to make the choice that will cause the least harm. This leaves important ethical questions of decision-making, responsibility, and of the worth of human life. Object detection and decision systems therefore should be developed in consideration of ethical rules that guarantee human safety, fairness and responsibilities as opposed to technology-focused optimization.

4.5 Standards

The proposed system follows established machine learning and neuromorphic computing standards, including best practices for model evaluation, reproducibility, and documentation. Standard object detection metrics such as precision, recall, and mean Average Precision (mAP) are considered for performance evaluation. The research also aligns with general software engineering standards for modular design, version control, and experimental reporting.

4.6 Project management plan

The project is carried out in accordance with a systematic timeline that is divided into several stages, such as literature review, model design, implementation, experimentation, and analysis. The first stages aim at familiarizing oneself with the RSNN network and spike-based learning. The next steps include the implementation of the model, preparation of datasets, and evaluation of the performance. The resources can be computing hardware, software tools, and academic references. Milestones and progress reviews are regular and made to ascertain that the project is completed in time.

The project was done in stages in order to secure systematic progress. Firstly, the literature associated with RSNN and energy-efficient object detection was conducted in order to establish the scope of the research. The second step was the design of the

RSNN-based architecture and choice of suitable datasets. It was then implemented and experimented on with the help of simulation tools and performance evaluation and analysis were done. Proper progress monitoring and periodic review of the progress were ensured to carry out the project in time.

4.7 Risk management

The proposed research has a number of risks. The process of training RSNNs can be computationally intensive and can be prone to convergence problems because of non-differentiable spike functions. In order to reduce this risk, surrogate gradient techniques and well established neuron models are used. The other risk is the fact that it may result in worse performance than traditional deep neural networks; this can be mitigated through parameter tuning and a combination of approaches. Widespread compatibility Hardware and software compatibility risks are reduced by utilizing popular frameworks.

There were a number of risks that were identified in the development of the project. A significant risk was that it was not easy to train RSNN models because of non-differentiable functions of spikes, which were reduced by adopting surrogate gradient methods. The other threat was that there was a possible gap in performance with the traditional deep neural networks, which was overcome by the careful parameter tuning and model optimization. Software or hardware constraints were the technical risks that were eliminated with the assistance of highly supported structures and backup settings.

4.8 Economic analysis

Economically, object detection systems that are based on the RSNN have long-term cost-saving as they use less energy and cost less to operate. The implementation of energy-efficient models reduces the cost of hardware and the consumption of power in the long run, even though the initial development and experimentation might need specialized knowledge and computation resources. This renders the proposed solution to be economically feasible to large scale and edge based intelligent systems, especially in low resource environments.

Chapter 5

Proposed Methodology

5.1 Design process or methodology overview

A hybrid Faster R-SNN is proposed in this study by expanding the conventional Faster R-CNN framework with spiking neuron dynamics which is influenced by biological neurons to detect objects in autonomous driving applications. The key motivation of this research is to present temporal and event-driven calculation into the pipeline of object detection, conversely handling the localization ability and the high detection accuracy of Faster R-CNN.

Faster R-CNN completely depends on artificial neural networks (ANNs) using continuous activations leading to high processing cost and power usage as a result of complex and recurrent operations. To overcome this limitation, the suggested Faster R-SNN is an addition of the spiking neural network (SNN) processing into the Faster R-CNN pipeline, in an ANN-SNN framework. The CNN backbone is kept the same and it generates regular feature maps which are then subjected to LIF neurons to transform them into the spike-based representations. This spike-enhanced feature map is inputted into the Region Proposal Network (RPN) and detection heads which is left to work in an analog mode. The model applies spiking dynamics without any changes to either the backbone or detection pipeline, which is achieved by connecting LIF neurons between feature extraction and the RPN.

Temporal spike accumulation and event-based processing of visual features are facilitated by the incorporation of Leaky Integrate and Fire (LIF) neurons. This enables the network to represent visual information throughout several discrete time steps that decreases redundant activity of neurons and enhances robustness in difficult conditions for example- occlusion and noise. Simultaneously, maintaining the analog RPN and detection heads assures the stability of the training and precise bounding box prediction.

The presented Faster R-SNN model's complete pipeline along with ANN modules, spiking modules and spike to spike transition is displayed in figure 5.1.

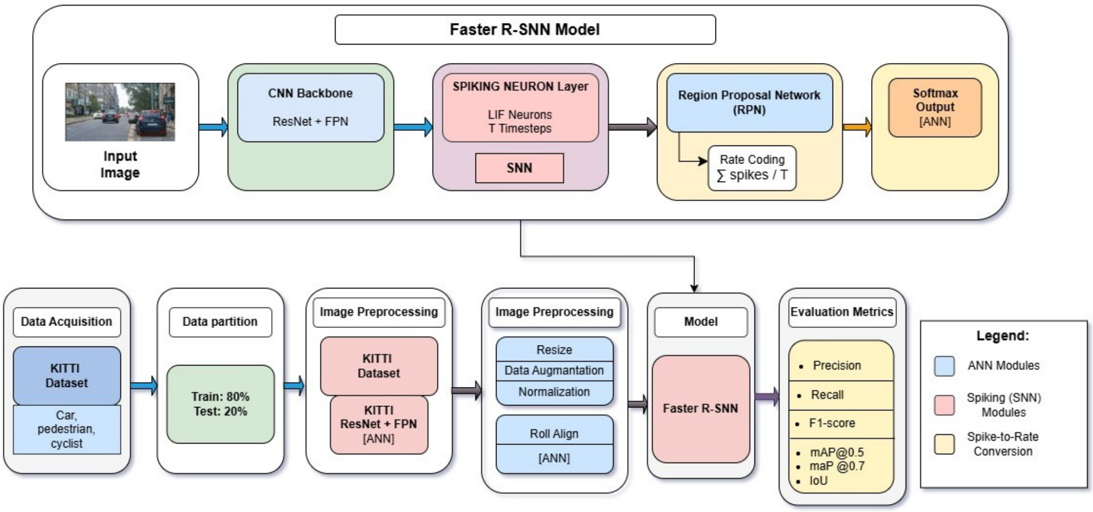

Figure 5.1: R-SNN Architecture Diagram

5.2 Leaky Integrate-and-Fire (LIF) Neuron Model

In the proposed Faster R-SNN, neurons are presented by Leaky Integrate-and-Fire (LIF) which are used to simulate spiking behavior. The following figure represents the internal functioning of the LIF neuron. The figure 5.2 illustrates the manner in which the neuron receives the input signals, generates spikes and restores its condition in the course of time.

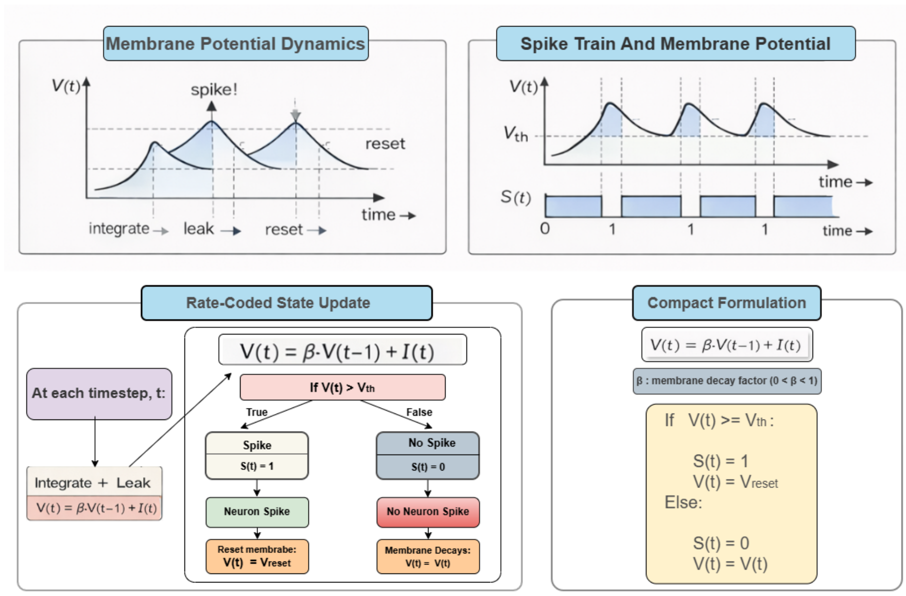

Figure 5.2: Internal Dynamics of LIF neuron illustrating membrane integration, spike generation and reset mechanism over time

The value of the membrane potential is a time-dependent value of each LIF neuron. At every time step t, the membrane potential update of V(t) by the following rule:

$$V(t) = \beta V(t - 1) + I(t)$$

Here, the LIF membrane update rule is a general discrete-time approximation of the classical leaky integrate-and-fire neuron model, which is commonly used as a spiking neural network simulation [30]. I(t) is the current input of the CNN feature maps and β is the leakage factor, which has a value ranging between 0 and 1. This leakage leads to the slow decrease of the membrane potential in case no new input arrives. Consequently, the neuron has the ability to retain the recent inputs and forget the old inputs which helps the network to process information with time rather than at once [31].

Once the membrane potential reaches a certain large value and exceeds a predetermined level Vth, the neuron produces a spike. It is a binary signal, i.e. a spike, which means that the neuron has been excited. The membrane potential returns to a preset value after discharging (Vreset). A failure to trigger the membrane potential to the threshold will result in the production of no spikes and the neuron will keep on integrating incoming input. This combination of integration, firing and reset enables the neuron to be time-based and event-driven.

The spike output S(t) represents the activation of a particular neuron at a given time period. The magnitude of the input signal is transferred into the amount of spikes a neuron produces within a sequence of steps. This is what is called rate coding [32]. The spiking neuron layer converts the temporal spike activity to a continuous value which may be usable by the next analog layers in the network by counting spikes over time. The technique helps the model to achieve the temporal dynamic of visual features and improve the consistency of object recognition.

5.3 Data preprocessing

The preprocessing stage prepares the raw KITTI dataset for utilizing it efficiently during training and testing. Initially, the image and annotation file paths are structured and object classes are allocated to the numerical labels. The annotation files are then ready to retrieve significant information such as object categories and bouncing box coordinates. To ensure data quality, invalid entries, unknown classes and incorrect bounding boxes are eliminated. Then all valid labels and bounding boxes are transformed into an organized numerical format which is suitable for deep learning models. Furthermore, to maintain evaluation and robust model training, the dataset is divided into 80% training and 20% testing.

Subsequently, a custom dataset loader is utilized to pair each image with its equivalent annotations. While the processed labels are stored in a standardized target structure, the images are loaded and transformed into a consistent RGB format. Supplementary information such as bounding box area and image identifiers is also incorporated for evaluation objectives. To enhance the model robustness, optional data transformations are employed. Finally, samples are integrated into batches utilizing a custom function. Through the procedure, the raw dataset is cleaned, standardized and converted into a model ready state for efficient object detection training.

5.4 Implementation of selected design

This model is trained and evaluated utilizing the benchmark KITTI object detection dataset which is broadly applied for the research of autonomous driving. The dataset consists of multiple labelled scenarios of real world streets with object categorization such as car, cyclist and pedestrian which are crucial for safety driven applications.

For training and testing, the dataset is split into two subsets. Between these two subsets, 80% is used for training and 20% is used for testing. Standardized preprocessing approaches including resizing and normalization are employed before inputting images into the network. The resolution of the input images are both consistent and stable throughout batches over the period of training which is ensured because of these steps.

5.4.1 CNN Backbone with Feature Pyramid Network (FPN)

The proposed Faster R-SNN model's backbone is based on the standard ResNet with Feature Pyramid Network (FPN) framework that is utilized in the Faster R-CNN architecture. The backbone of the CNN derives the representations of hierarchical features at multiple spatial scales which permits detection of both small and large objects.

This backbone, in terms of structure, is maintained to sustain the deep convolutional networks' strong representational power in this hybrid model. The multi scale feature maps are produced by the FPN function as input signals to spiking neuron layers which facilitate a seamless transition from continuous feature representations to spike based processing.

5.4.2 Spiking Neuron Layer

The primary contribution of this proposed hybrid model is embodied in the implementation of a spiking neuron layer directly after the CNN backbone. This layer substitutes conventional activation functions with LIF neurons.

Each neuron implements input feature activations across T discrete time steps where its membrane potential progresses as per LIF dynamics. A spike is produced and the potential is reinitialized when the membrane potential surpasses a preassigned threshold. This procedure transforms static feature maps into temporal spike trains which allows the network to collect evidence throughout time instead of depending on a single forward pass.

This model does not use feedback or iterative connections different from recurrent neural networks. Rather, temporal information is obtained via spike integration throughout time steps which delivers temporal robustness excluding incorporating architectural complexity.

5.4.3 Spike to Rate Coding Conversion

A spike to rate coding method is applied to interface spiking and non spiking elements since the RPN and detection perform in the analog area. The binary spike outputs produced by the LIF neurons are aggregated throughout T time steps and calculated as the mean to configure continuous feature maps which is shown in the following through equation:

$$\text{Rate} = \frac{1}{T}\sum_{t=1}^{T} S(t)$$

Here, S(t) denotes the binary spike output at t time step.

This rate coded representation maintains the temporal information mapped by spiking activity, concurrently remaining consistent with standard convolutional and regression analysis [32]. Therefore, the hybrid model integrates SNN and ANN modules effectively without altering the pipeline of Faster R-CNN.

5.4.4 Region Proposal Network (RPN) and Detection Head

The RPN obtains the rate coded feature maps and produces potential object proposals by estimating scores of the probability of being an object and the offsets of the bounding box. This level is maintained fully analog to secure precise localization and stable optimization.

Upon completion of the RPN, RolAlign is employed to retrieve fixed size feature representations for each proposal. These features are forwarded to the detection head which carries out object classification utilizing a layer of softmax and enhances bounding box coordinates [27].

The proposed model prevents instability linked to spiking regression layers and attains detection performance equivalent to conventional Faster R-CNN while taking advantage of spiking computation in the backbone by managing these phases in the analog domain.

Chapter 6

Results

6.1 Initial Analysis

In our experiments on the KITTI dataset, before the proposal of the hybrid model, we initially evaluated both spiking-based and conventional deep learning object detection models and observed clear performance differences. The SNN-based models, namely Spiking-YOLO and SUHD, showed comparatively lower detection accuracy. Spiking-YOLO achieved an overall precision of 29.91%, recall of 17.21%, and mAP@0.5 of approximately 0.05, while SUHD obtained a precision of about 19.8% and recall of 30.1%. Although both spiking models were able to detect larger and more frequent object classes such as cars and trucks, they consistently struggled with smaller or less frequent classes like pedestrians and cyclists, resulting in low recall and higher false-negative rates.

In contrast, the DNN-based models demonstrated substantially stronger performance across all evaluation metrics. YOLOv5 achieved the best overall results, with a micro-average precision of 82.48%, recall of 90.39%, and F1 score of 86.25%, indicating robust and reliable object detection across all classes. The SSD model also performed competitively, achieving a precision of 0.831, recall of 0.439, and mAP@0.5 of 0.423, offering a balanced trade-off between accuracy and computational efficiency. Overall, these results indicate that while SNN-based detectors show promise due to their potential energy efficiency, their detection performance currently lags behind conventional DNN-based models, emphasizing the need for further architectural and training improvements in spiking networks.

After seeing these outcomes, we developed and evaluated the proposed hybrid model.

6.2 Performance evaluation

Disclaimer: All experiments and evaluations presented in this study were conducted in a controlled simulation environment using Jupyter Notebook with the PyTorch deep learning framework on a desktop system running Windows 11 and equipped with an NVIDIA GeForce RTX 3070 Ti graphics card. GPU power consumption was measured using MSI Afterburner, which provides software-based estimates of GPU power usage during training and inference. As both performance and energy measurements are influenced by the underlying hardware, operating system, software framework, driver behavior, and monitoring methodology, the reported results are specific to this experimental setup and should be interpreted within this context. Consequently, these findings are not claimed to be universally generalizable across different platforms, configurations, or real-world and neuromorphic deployment environments, and further evaluation under diverse conditions would be required to establish broader generalizability.

The Faster R-CNN model and Hybrid model's performance is analyzed by using the benchmark KITTI dataset which consists of eight types of object classes such as car, pedestrian, truck, van, person sitting, cyclist, tram and miscellaneous. The dataset is segmented as training and testing subsets that follow established evaluation protocols.

Object detection accuracy is mainly evaluated utilizing Intersection over Union (IoU) where a predicted bounding box is regarded as correct if its overlap with the ground truth is greater than a pre-specified threshold. In this research, an IoU threshold of 0.5 is applied as the baseline standard and the corresponding mean Average Precision (mAP@0.5) is utilized as the main evaluation metric because of its broad adoption in autonomous driving standards. To analyze further, localization precision under more stringent conditions, additional evaluations are performed utilizing mAP@0.7 and COCO style mAP@[0.50:0.95] which averages performance throughout multiple thresholds of IoU and delivers a more in-depth measure of detection robustness.

Additionally, detailed class-wise evaluation is conducted using Average Precision (AP) for each object class to evaluate class specific behaviour along with global mAP values. Precision, recall, F1 score metrics are further integrated to measure the trade-off between false positives and false negatives which is specifically crucial for safety critical autonomous driving applications. Furthermore, to evaluate misclassification patterns among visually similar object categories, confusion matrices and related visualizations are used which facilitates deeper understanding into model strengths and limitations.

6.2.1 Performance evaluation workflow of Hybrid R-SNN

The presented hybrid architecture's overall performance for the detection on the KITTI object detection dataset is condensed in the following manner:

- mAP@0.5: 0.7395
- mAP@0.7: 0.5926
- mAP@[0.50:0.95]: 0.4771

The high value of mAP@0.5 shows the model's high detection efficacy based on conventional evaluation criteria. Furthermore, the values of mAP@0.7 and mAP@[0.50:0.95] indicate precise localization performance utilizing higher overlap thresholds. These outcomes validate that this model can not only localize but also classify objects throughout various levels of spatial precision efficiently.

Per Class Quantitative Analysis:

The trained hybrid model acquired class wise precision, recall, F1 score, truth positive (TP), false positive (FP) and false negative (FN) are calculated on the basis of the final results of the evaluation. The numerical results for the object classes are presented below:

Class	Precision	Recall	F1 Score	Support	TP	FP	FN
Car	0.4148	0.9900	0.5848	5843	5673	8002	170
Cyclist	0.1652	0.9493	0.2814	283	263	1329	20
Misc	0.0839	0.8623	0.1528	184	155	1692	29
Pedestrian	0.1541	0.9194	0.2642	935	841	4615	94
Person_sitting	0.1105	0.6775	0.1900	73	48	386	25
Tram	0.1771	0.9311	0.2976	90	82	381	8
Truck	0.2660	0.9779	0.4183	238	228	629	10
Van	0.2332	0.9695	0.3761	575	546	1795	29

Table 6.1: Precision, Recall, F1 Score, Support, true positive (TP), false positive (FP) and false negative (FN) for each class.

From the outcomes, it is observed that though across most classes recalls are high which shows the capability of the spiking based temporal processing to identify a large portion of ground truth objects, precisions of several classes are relatively lower because of false positives, specifically in visually similar categories. This compromise indicates that the model emphasizes object detection coverage which is beneficial for safety essential systems like autonomous driving.

Additionally, class wise AP (Average Precision) is calculated to study the performance of the detection throughout multiple categories of object. The outcomes are given in a table in the following:

Class	Average Precision (AP@0.5)
Truck	0.9093
Car	0.9379
Van	0.8680
Tram	0.8036
Cyclist	0.8180
Misc	0.6109
Pedestrian	0.7443
Person_sitting	0.2234

Table 6.2: Average Precision (AP@0.5) for each class

Here, the results demonstrate that the model accomplishes very high AP for primary classes such as car, truck and van which are neatly presented in the KITTI dataset whereas the lower AP scores are detected for visually ambiguous classes such as person sitting which illustrates the challenge of detecting objects with low frequency.

Height based Average Precision Analysis:
To analyze the performance of the detection under varying levels of object difficulty, height based Average Precision AP@0.5 values are calculated for the dominant classes such as car, pedestrian and cyclist that follow the KITTI benchmark criteria.

Difficulty	Car	Pedestrian	Cyclist
Easy	0.9454	0.8313	0.7979
Moderate	0.9110	0.4515	0.8253
Hard	0.9329	0.1165	0.6675
Average	0.9298	0.4664	0.7636

Table 6.3: Difficulty-wise Average Precision (AP@0.5)

These outcomes demonstrate that the model performs significantly for easy and moderate objects especially for cars and cyclists. For hard pedestrian cases, performance gets lower which generally include small sizes of objects, occlusions and images with low resolution.

Qualitative and Visual Analysis:

To supplement the quantitative outcomes and to illustrate the learning behaviour and detection ability of the presented hybrid spiking deep neural network, qualitative and visual representations are displayed here.

Training Dynamics:

The training procedure is evaluated utilizing loss versus epochs and mAP@0.5 versus epoch graphs. Figure 6.1 and 6.2 represents the graphs respectively.

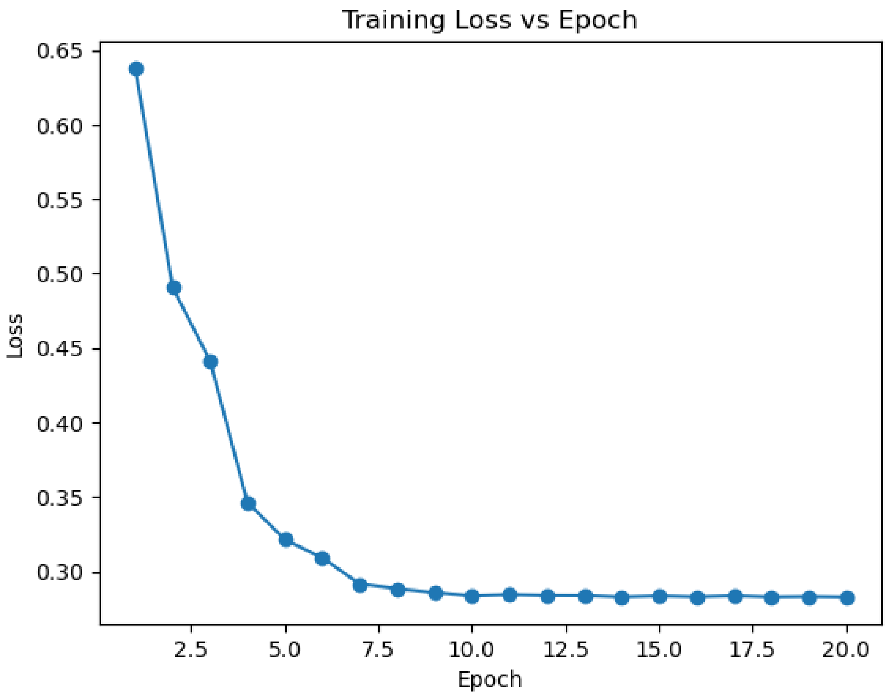

Figure 6.1: Training Loss Vs Epoch (Hybrid Model)

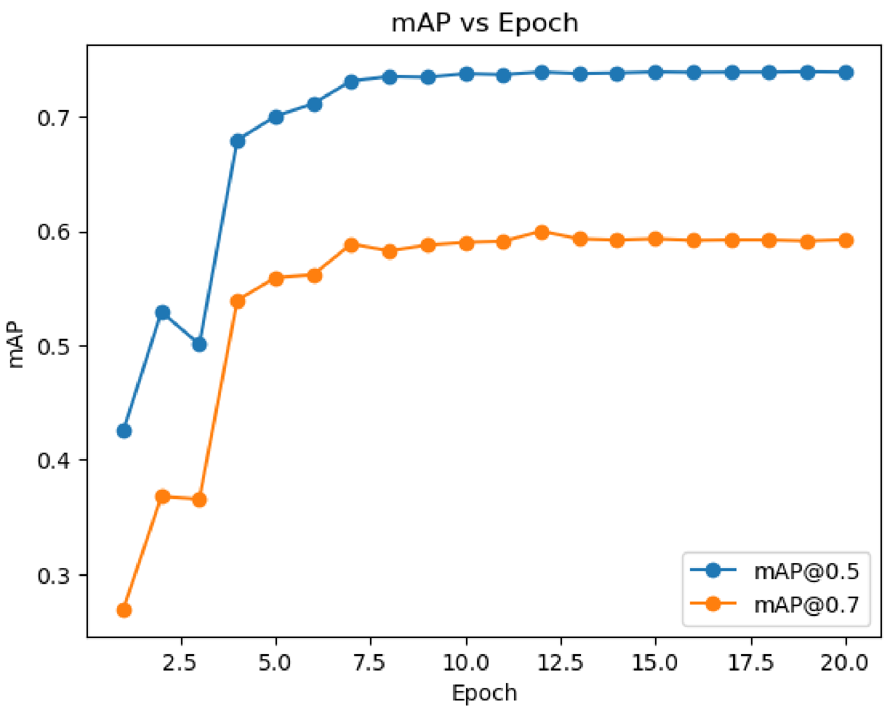

Figure 6.2: mAP vs Epoch (Hybrid Model)

The training loss demonstrates that there has been a steady reduction while the mAP@0.5 curve enhances and stabilizes over epochs that verifies proper learning and convergence of the model.

Quantitative Performance Visualizations:

Various graphical evaluations are utilized to describe the performance of the detection:

- Precision-Recall curves are illustrated for car, pedestrian and cyclist classes that show the balance between precision and recall throughout the thresholds of confidence. The dotted curve in Figure 6.3 illustrates the performance of the overall Precision-Recall of the model by merging detections from the classes of all objects.

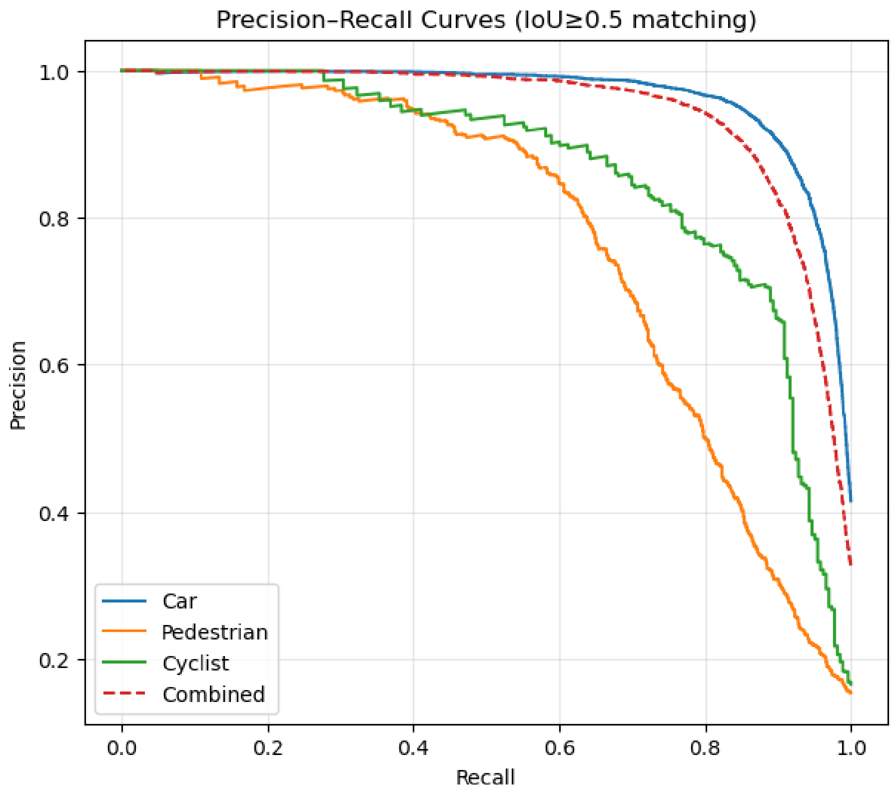

Figure 6.3: Precision-Recall Curves (Hybrid Model)

- The confusion matrix in Figure 6.4 demonstrates misclassification patterns, especially between visually similar classes.

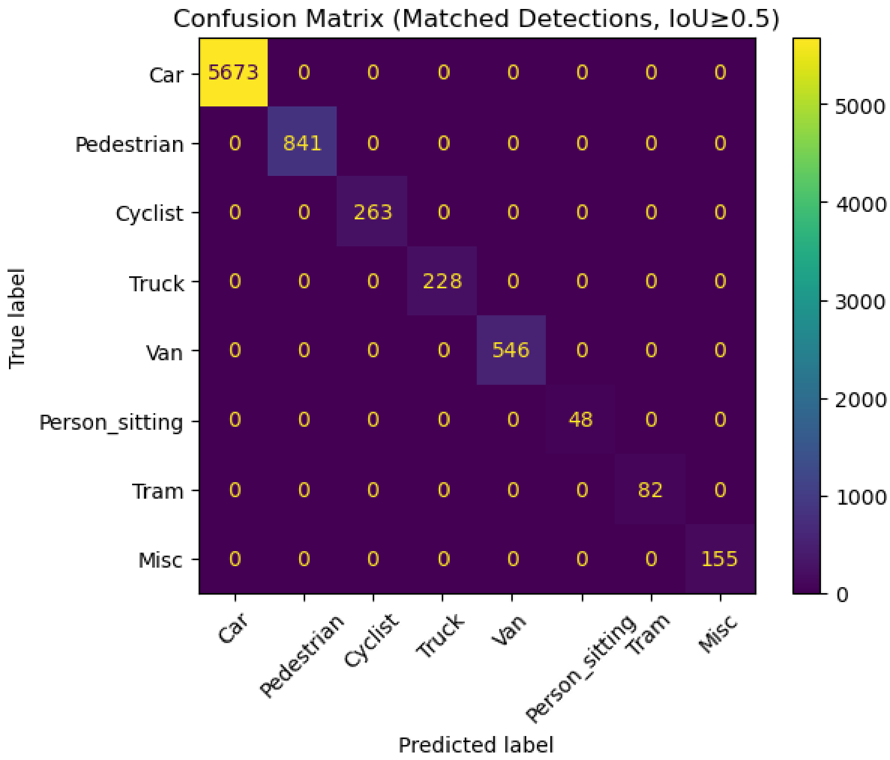

Figure 6.4: Confusion Matrix (Hybrid Model)

- The false negatives are more recurring for smaller and harder object classes which is demonstrated by a false positive vs false negative bar chart in Figure 6.5.

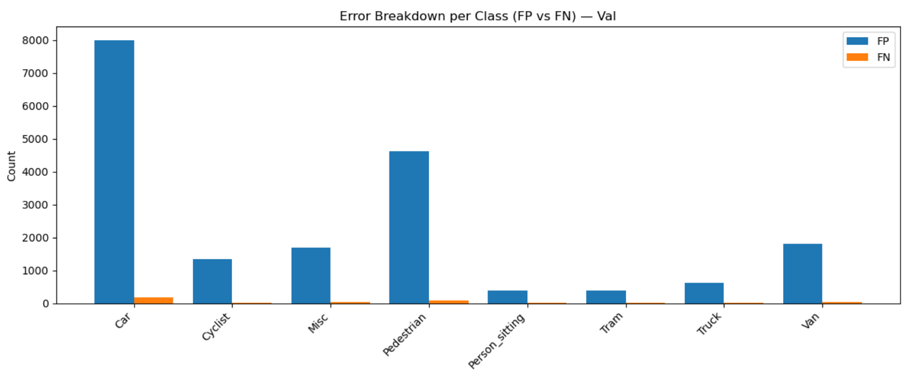

Figure 6.5: False Positive vs False Negative Bar Chart (Hybrid Model)

Qualitative Detection Results:

Sample detection images along with bounding boxes, class labels and confidence scores are displayed in Figure 6.6 to qualitatively verify the accuracy of detection through visual inspection

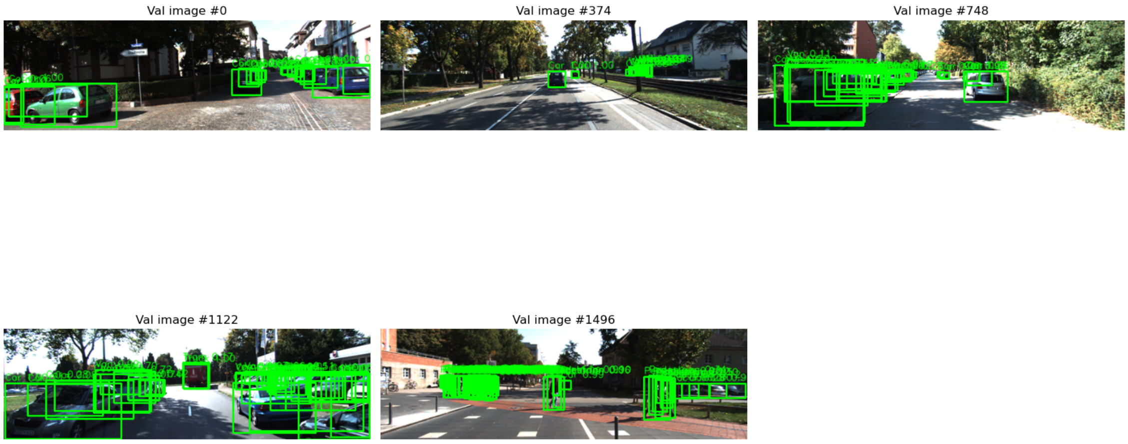

Figure 6.6: Sample detection results on KITTI images demonstrating predicted bounding boxes with class labels and confidence scores (Hybrid Model)

Moreover, to achieve a clearer understanding of dataset difficulty, the ground truth bounding box size distribution is evaluated. The distribution in visualized in Figure 6.7 below.

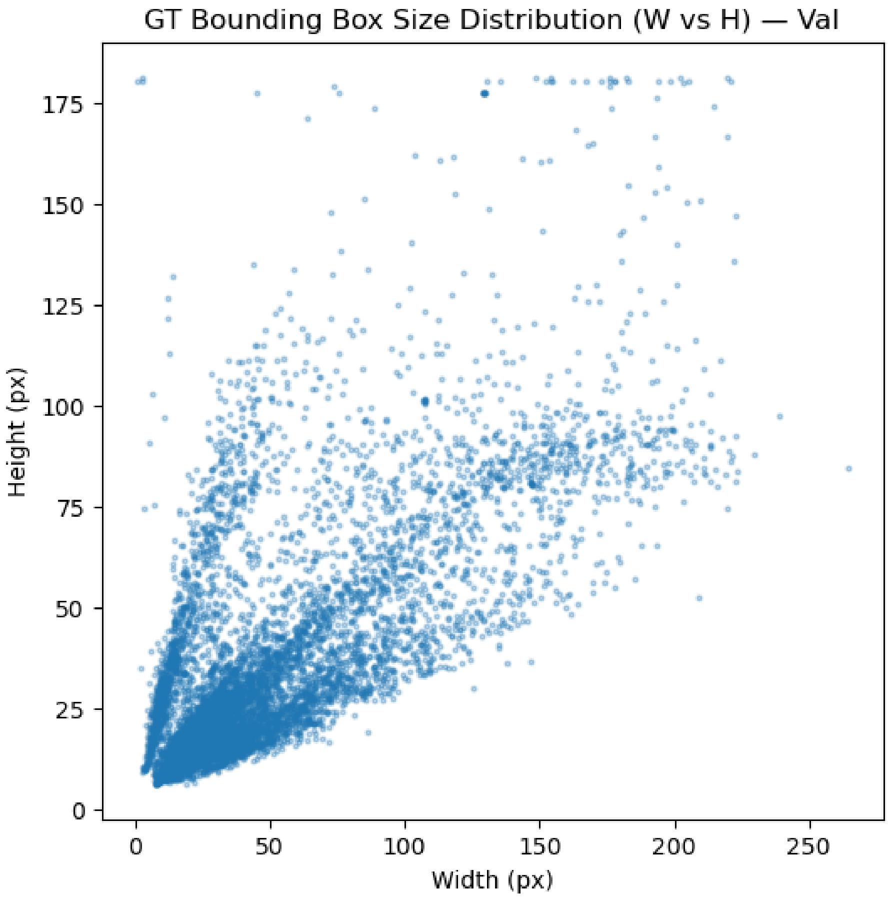

Figure 6.7: Ground truth bounding box size distribution for the KITTI validation set (Hybrid Model)

In Figure 6.8, TP(green), FP(red) and FN(blue) overlays evidently to demonstrate accurate detections, inaccurate detections and objects that are missed. A confidence heatmap is incorporated to display regions of the activation of high spiking neurons that shows efficient event based feature localization as shown in Figure 6.9.

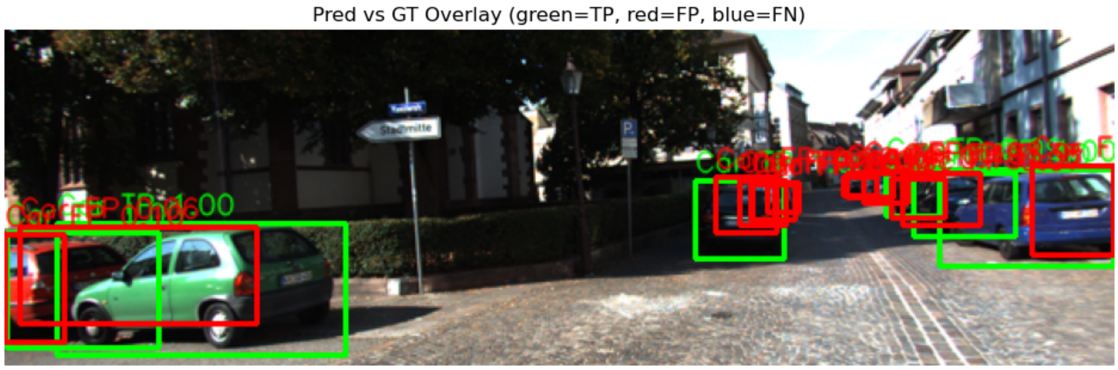

Figure 6.8: Visualization of true positives (TP), false positives (FP) and false negatives (FN) in Hybrid Model results of detection

Figure 6.9: Confidence heatmap of detection predictions (Hybrid Model)

Analytical Visual Insights:

The IoU distribution histogram in Figure 6.10 demonstrates that most predictions acquire high convergence with ground truth boxes which validates correct localization.

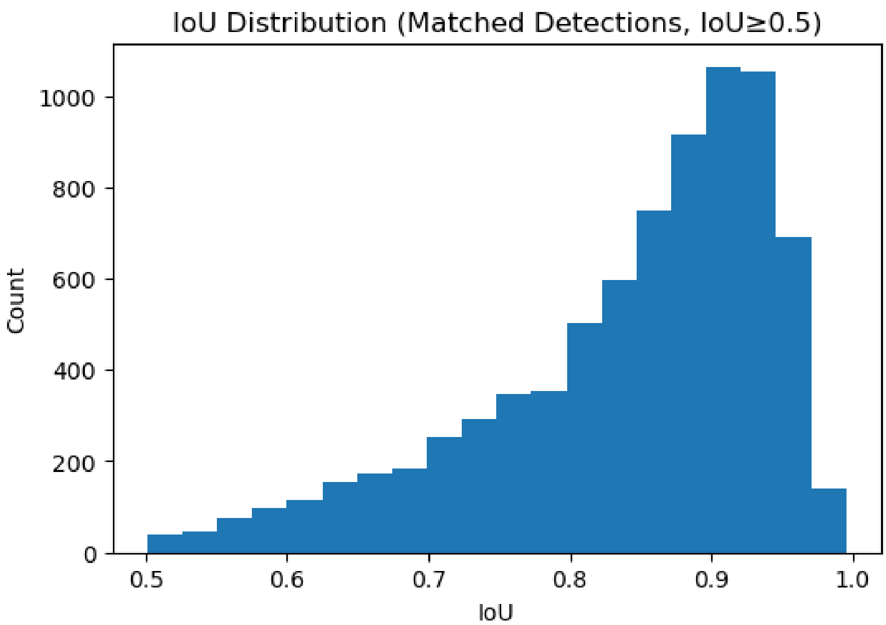

Figure 6.10: Histogram of IoU distribution (Hybrid Model)

Power Consumption Logging:

The following graph in Figure 6.11 shows the value of sustained power consumption while running the model (in Watts).

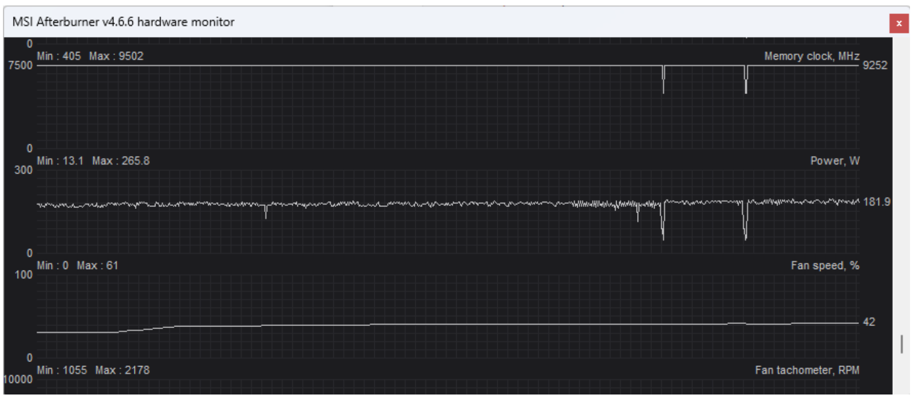

Figure 6.11: Power Consumption (Hybrid Model)

6.2.2 Performance evaluation workflow of Faster R-CNN

The summary of the overall performance of the Faster R-CNN architecture on the KITTI dataset on object detection is presented in the table below. The model also showed significant progress during the training process reaching high precision in localization and classification.

- mAP@0.5: 0.8590
- mAP@0.7: 0.7358
- mAP@[0.50:0.95]: 0.5920

Results show the model converged full of success and the mAP@0.5 attains its maximum of 0.8590 on the last analysis. This high score is accompanied by a high mAP 0.7 of 0.7358, proving the fact that the model not only identifies objects but also their location with a high level of overlap with the ground truth.

Per Class Quantitative Analysis:

An evaluation of the performance of the model was done in detail per class to determine the reliability of the model in the various object classes in the KITTI dataset. The next table shows the accuracy and recall as well as the F1-score of each class and the exact numbers of False Positives (FP) and False Negatives (FN).

Class	Precision	Recall	F1-Score	Support	TP	FP	FN
Car	0.5761	0.9706	0.7231	5843	5671	4172	172
Cyclist	0.2675	0.9435	0.4169	283	267	731	16
Misc	0.2175	0.9185	0.3517	184	169	608	15
Pedestrian	0.2762	0.9102	0.4238	935	851	2230	84
Person_sitting	0.1921	0.7945	0.3093	73	58	244	15
Tram	0.3857	0.9556	0.5495	90	86	137	4
Truck	0.4726	0.9790	0.6375	238	233	260	5
Van	0.4107	0.9600	0.5753	575	552	792	23

Table 6.4: Precision, Recall, F1 Score, Support, true positive (TP), false positive (FP) and false negative (FN) for each class.

The findings indicate that the model works exceptionally well with the most frequent object in the data set which is the Car category. Its overall performance was good with an F1-score of 0.7231. Most of the categories also have very high recall values in the model with the highest being 0.9706 which is Car and 0.9790 which is Truck. This implies that the Faster R-CNN model is effective in detecting objects and it does not miss objects easily. That is, it has a very low false negative.

Nevertheless, other categories get a low score of precision, particularly Person_sitting and Misc, with the values of 0.1921 and 0.2175 respectively. This implies that although the model identifies the majority of these objects, it is highly prone to errors by identifying objects that do not exist. Indeed, in the Person_sitting category, the model had 244 false positives but 58 true positives. It indicates that the model is identifying numerous false objects as well as the true ones.

Such an issue is typical to object detection, in particular, small or less common classes. The model is then over sensitive and will start recognizing background patterns or other similar forms as real objects. Although it does not lose many objects, it requires an improved accuracy to minimize these false detections.

Additionally, class wise AP (Average Precision) is calculated to study the performance of the detection throughout multiple categories of object. The outcomes are given in a table in the following:

Class	Average Precision (AP@0.5)
Truck	0.9583
Car	0.9513
Van	0.9285
Tram	0.9049
Cyclist	0.8834
Misc	0.8325
Pedestrian	0.8228
Person_sitting	0.5907

Table 6.5: Average Precision (AP@0.5) for each class

The findings confirm that the model has high AP values over prevailing classes of vehicles, including Car (95.13%), Truck (95.83%), and Van (92.85%), which have a high presence in the KITTI dataset. On the other hand, less frequent or more complex visually, the classes (Person-sitting 59.07%), got lower AP scores. This gap makes it clear why it is intrinsically difficult to identify objects with large pose variations or fewer training examples where they can be easily confused with the more prevalent Pedestrian category.

Height based Average Precision Analysis:

To further examine the strengths of the proposed model, an assessment was done depending on the height of the pixel bounding boxes of the objects. Object height is among the major aspects of the KITTI benchmark since it is highly associated with the distance between the object and the sensor and directly influences the difficulty of detection. According to the dimensions of vertical pixels, three levels of objects are illustrated: Easy (large objects), Moderate (medium-sized objects), and Hard (small objects).

Difficulty	Car	Pedestrian	Cyclist
Easy	0.9677	0.8829	0.9061
Moderate	0.9572	0.5810	0.9279
Hard	0.9567	0.2188	0.8000
Average	0.9605	0.5609	0.8780

Table 6.6: Difficulty-wise Average Precision (AP@0.5)

The Faster R-CNN that uses a Feature Pyramid Network (FPN) shows a stable performance in all these levels of difficulty. For large objects, the model has high detection rates as demonstrated in the difficult-wise AP@0.5 results as the model has high scores in Average Precision of Cars (0.9677), Pedestrians (0.8829), and Cyclists (0.9061) in the Easy category. It means that the model is effective to address high spatial and semantic features in the image when the objects are represented by a larger number of pixels. Medium-sized objects do not decrease performance, and the Car (0.9572) and Cyclist (0.9279) classes in particular seem to be advantageous in the case of the multi-scale feature representation offered by the FPN.

Nonetheless, performance of detection reduces on small objects in the Hard group, particularly on Pedestrian class that reduces to an AP of 0.2188. This drop is indicative of the famous problem related to the ability of the region-based detectors to detect distant and low-resolution objects. Smaller objects occupy fewer pixels and it becomes difficult to get discriminative features out of such a model. Nonetheless, the Car and Cyclist classes have relatively high AP values of 0.9567 and 0.8000 respectively, as they have more unique shapes and visual shapes.

Qualitative and Visual Analysis:

To supplement the quantitative values, qualitative assessment was conducted through visualization of the Faster R-CNN detection results, in the KITTI test sequences. This discussion is based on the model as to whether it is capable of properly localizing objects, the confidence of the result, and the presence of overlapping or hidden cases.

Training Dynamics:

The training procedure is evaluated utilizing loss versus epochs and mAP@0.5 versus epoch graphs. Figure 6.12 and 6.13 represents the graphs respectively.

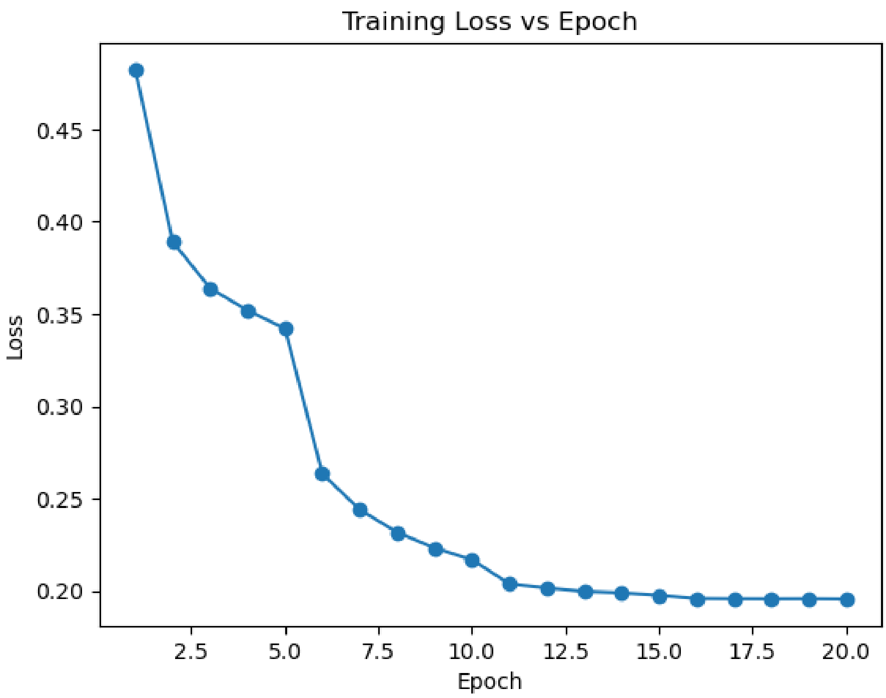

Figure 6.12: Training Loss Vs Epoch (Faster R-CNN)

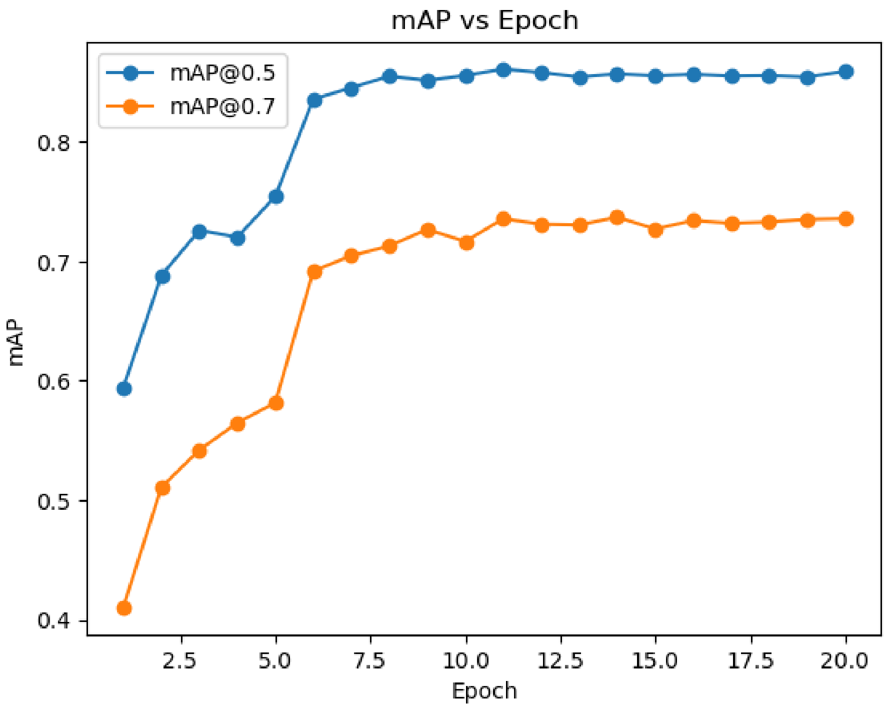

Figure 6.13: mAP vs Epoch (Faster R-CNN)

Quantitative Performance Visualizations:

Various graphical evaluations are utilized to describe the performance of the detection:

- Precision-Recall curves are illustrated for car, pedestrian and cyclist classes that show the balance between precision and recall throughout the thresholds of confidence. The dotted curve in Figure 6.14 illustrates the performance of the overall Precision-Recall of the model by merging detections from the classes of all objects.

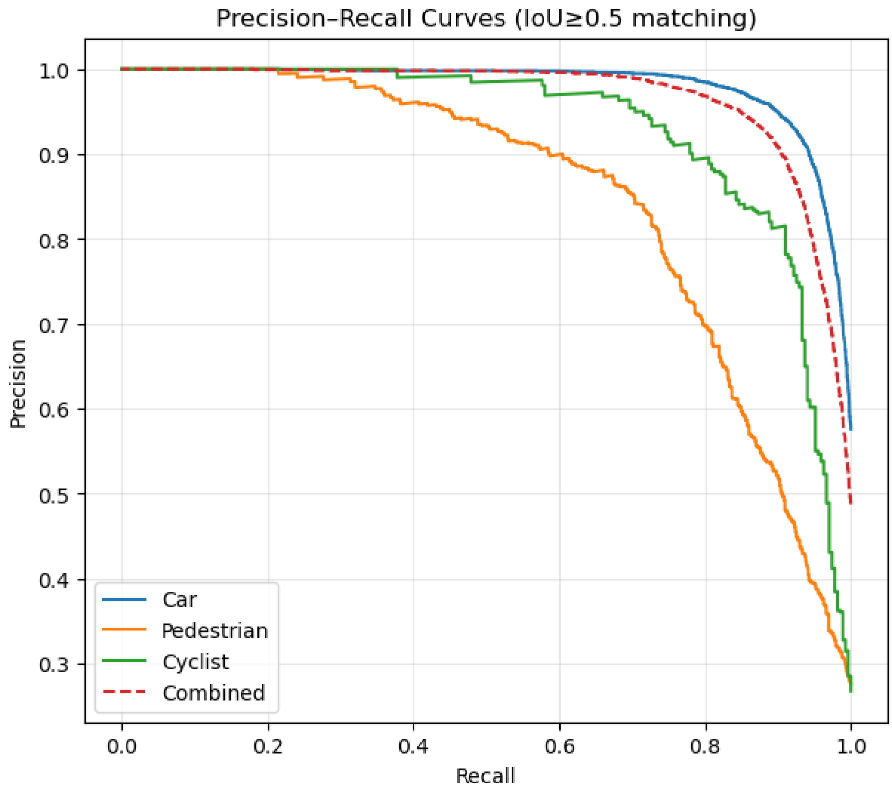

Figure 6.14: Precision-Recall Curves (Faster R-CNN)

- The confusion matrix in Figure 6.15 demonstrates misclassification patterns, especially between visually similar classes.

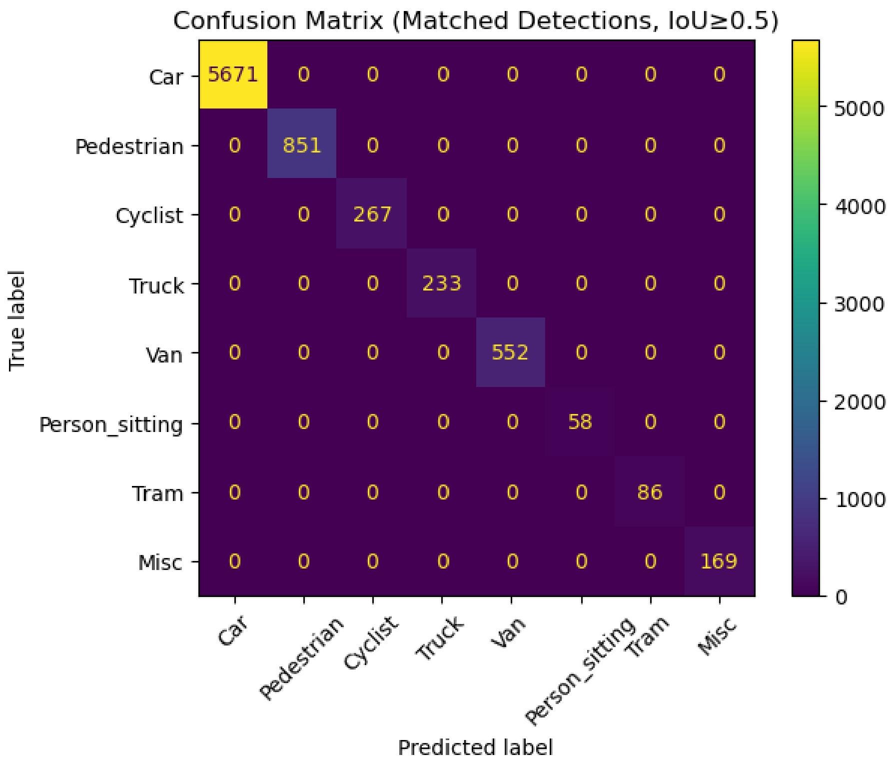

Figure 6.15: Confusion Matrix (Faster R-CNN)

- The false negatives are more recurring for smaller and harder object classes which is demonstrated by a false positive vs false negative bar chart in Figure 6.16.

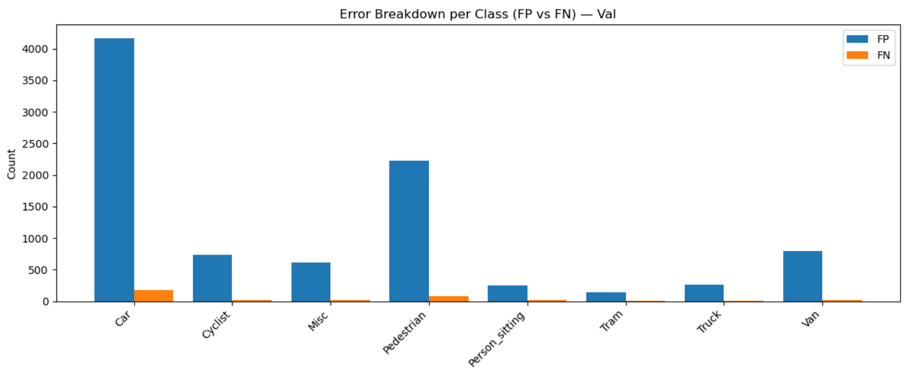

Figure 6.16: False Positive vs False Negative Bar Chart (Faster R-CNN)

Qualitative Detection Results:

Sample detection images along with bounding boxes, class labels and confidence scores are displayed in Figure 6.17 to qualitatively verify the accuracy of detection through visual inspection

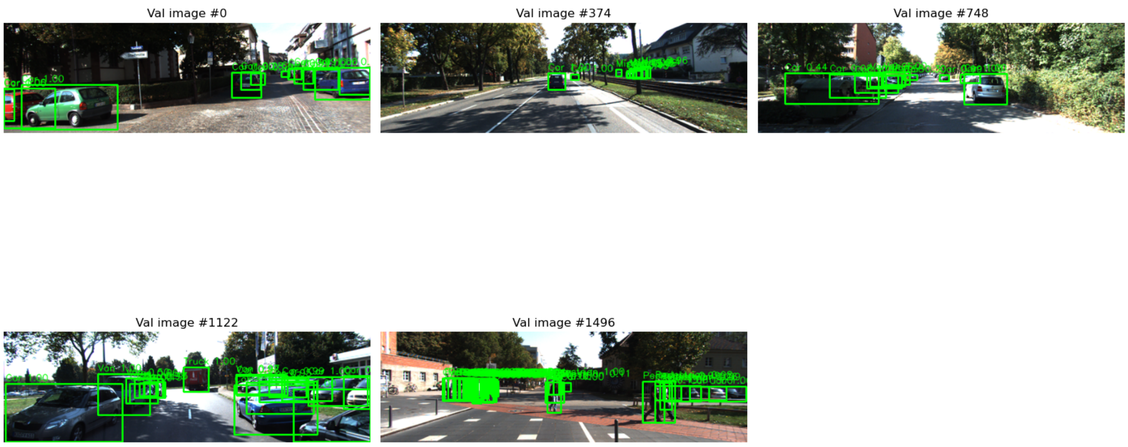

Figure 6.17: Sample detection results on KITTI images demonstrating predicted bounding boxes with class labels and confidence scores (Faster R-CNN)

Moreover, to achieve a clearer understanding of dataset difficulty, the ground truth bounding box size distribution is evaluated. The distribution in visualized in Figure 6.18 below

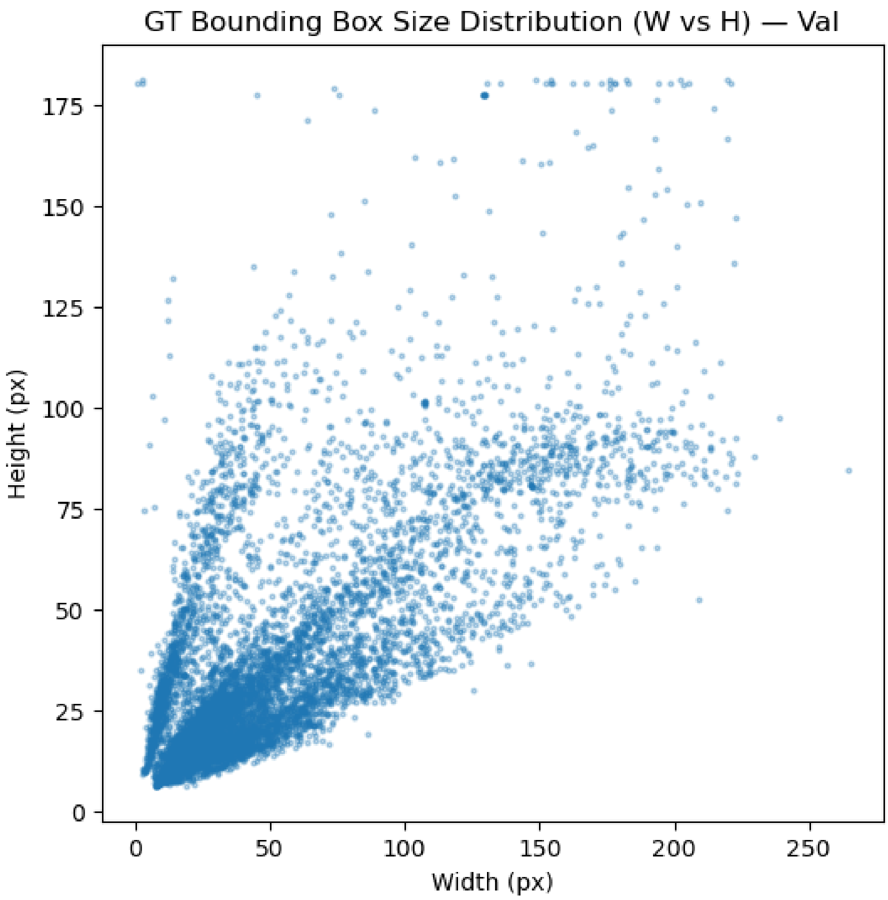

Figure 6.18: Ground truth bounding box size distribution for the KITTI validation set (Faster R-CNN)

In Figure 6.19, TP(green), FP(red) and FN(blue) overlays evidently to demonstrate accurate detections, inaccurate detections and objects that are missed. A confidence heatmap is incorporated to display regions of the activation of high spiking neurons that shows efficient event based feature localization as shown in Figure 6.20.

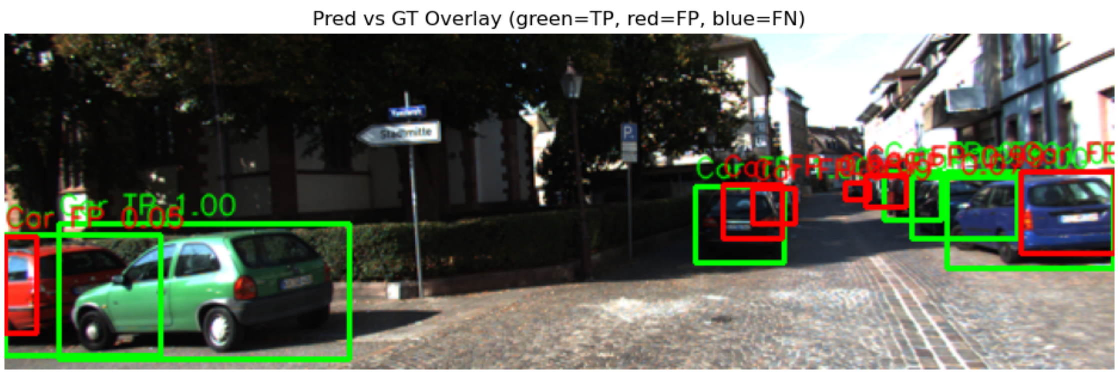

Figure 6.19: Visualization of true positives (TP), false positives (FP) and false negatives (FN) in Faster R-CNN results of detection

Figure 6.20: Confidence heatmap of detection predictions (Faster R-CNN)

Analytical Visual Insights:

The IoU distribution histogram in Figure 6.21 demonstrates that most predictions acquire high convergence with ground truth boxes which validates correct localization.

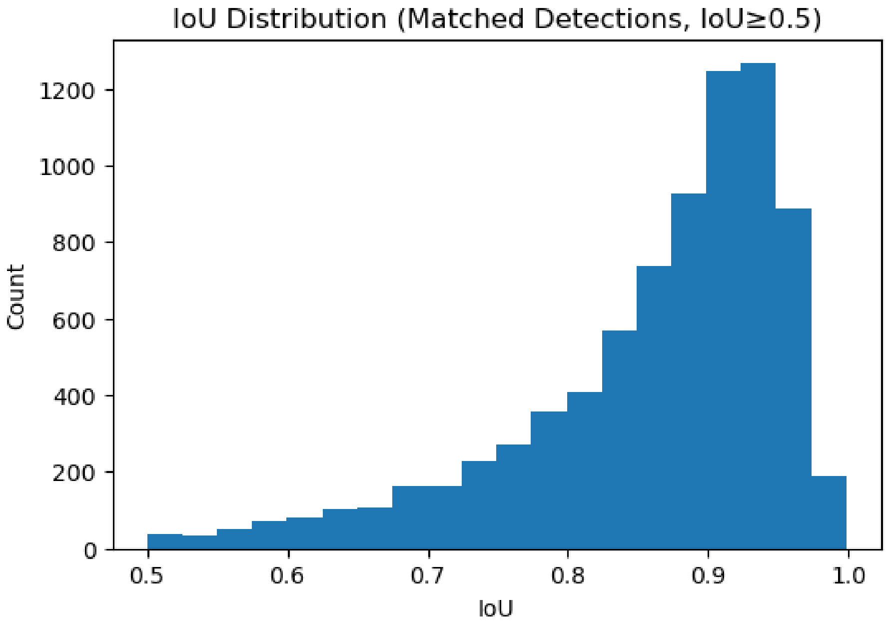

Figure 6.21: Histogram of IoU distribution (Faster R-CNN)

Power Consumption Logging:

The following graph in Figure 6.22 shows the value of sustained power consumption while running the model (in Watts).

Figure 6.22: Power Consumption (Faster R-CNN)

6.3 Final design adjustments

As part of the final design optimization, the training strategy of the proposed Hybrid R-SNN model was refined through careful selection of optimization and learning rate scheduling techniques. Stochastic Gradient Descent (SGD) was adopted as the primary optimizer, with a learning rate of 0.005 to ensure stable and gradual convergence. Momentum was set to 0.9 to accelerate training by smoothing gradient updates and reducing oscillations during optimization.

To control overfitting and improve generalization, L2 regularization was incorporated using a weight decay factor of 0.0005. Additionally, a Step Learning Rate Scheduler was employed, where the learning rate was reduced by a factor of 0.1 every three epochs. This adaptive scheduling strategy allowed the model to learn rapidly in early stages while enabling finer weight adjustments in later epochs.

These final training adjustments significantly improved convergence stability, reduced training loss, and enhanced the overall detection performance of the RSNN model.

6.4 Statistical analysis

The autonomous perception systems lack a generally standardized equation that can be used in determining the computational or energy efficiency. It is however common to measure object detection models on accuracy-latency trade-offs and even energy-sensitive performance, particularly in a real-time and edge-deployed setting where researchers benchmark object detectors based on mean average precision and inference time and associated performance estimates [33]. Moreover, energy- conscience deep learning research indicates the relevance of quantifying energy usage by real power and execution time and mapping these values on model effectiveness through indexes of energy efficiency [34], [35]. Based on these practices, we come up with the following task specific measures of efficiency.

Basic Efficiency

Basic efficiency is defined as the ratio of detection accuracy to total inference time, indicating how efficiently a model produces accurate detections over time.

Hybrid R-SNN

Basic Efficiency
= mAP@0.5 ÷ Inference Time
= 0.7395 ÷ 656.5008 Secs
= 0.00112 mAP/sec

Faster R-CNN

Basic Efficiency
= mAP@0.5 ÷ Inference Time
= 0.8609 ÷ 580.7490 Secs
= 0.00148 mAP/sec

The Faster R-CNN model demonstrates higher basic efficiency due to superior detection accuracy and reduced inference time, whereas the Hybrid R-SNN exhibits a modest reduction in speed and accuracy as a trade-off for energy efficiency.

Power Efficiency

Power efficiency quantifies the amount of detection accuracy achieved per unit of energy consumption.

Hybrid R-SNN

Energy
= Power × Execution Time
= 180 W × 13197.9239 Secs
= 2375626.302 Joules

Power Efficiency
= mAP@0.5 ÷ Energy
= 0.7395 ÷ 2375626.302
= 3.11E-7 mAP/J

Faster R-CNN

Energy
= Power × Execution Time
= 255 W × 11682.3536 Secs
= 2979000.168 Joules

Power Efficiency
= mAP@0.5 ÷ Energy
= 0.8609 ÷ 2979000.168
= 2.88E-7 mAP/J

Although the Hybrid R-SNN consumes lower total energy, it achieves higher power efficiency, highlighting the benefit of spiking neural computation in reducing redundant operations and enhancing energy-aware inference.

Multi-Metric Efficiency

Multi-metric efficiency incorporates both mAP@0.5 and mAP@0.7 to jointly assess detection accuracy and localization quality under time constraints.

Hybrid R-SNN

Multi-Metric Efficiency
= (mAP@0.5 + mAP@0.7) ÷ (2 × Inference Time)
= (0.7395 + 0.5926) ÷ (2 × 656.5008 Secs)
= 0.00101

Faster R-CNN

Multi-Metric Efficiency
= (mAP@0.5 + mAP@0.7) ÷ (2 × Inference Time)
= (0.8609 + 0.7357) ÷ (2 × 580.7490 Secs)
= 0.00137

The Faster R-CNN achieves higher multi-metric efficiency, indicating stronger performance across multiple accuracy thresholds. However, the Hybrid R-SNN maintains competitive performance while offering improved energy efficiency, making it more suitable for power-constrained edge applications.

6.5 Discussions

This discussion expounds the comparison of traditional Faster R-CNN model and the proposed Hybrid R-SNN model and demonstrates the merits of hybrid spiking-based object detection. As Faster R-CNN has a better detection accuracy within the typically used evaluation measures, the Hybrid R-SNN displays characteristics that will be part of the future energy-efficient and safety-critical vision systems.

The experimental results indicate that Faster R-CNN is better compared to Hybrid R-SNN in all measures of mean Average Precision (mAP) such as mAP at 0.5, mAP at 0.7 and COCO-like mAP at [0.50:0.95]. This proves its high capability of localizing objects and making accurate predictions. Nevertheless, the Hybrid R-SNN has a competitive detection performance when using sparse and event-based spiking computation. This shows that the spiking-based processing would be able to aid practical object detection requiring much lower computation activity.

Notably, the Hybrid R-SNN demonstrates an evident increase in accuracy compared to fully spiking object detection models tested previously in this paper. Earlier SNN based detectors like Spiking-YOLO and SUHD had very low mAP values and could not detect small or less frequently occurring classes of objects, e.g. pedestrians and cyclists. Conversely, the Hybrid R-SNN offers significantly better detection accuracy and recall between categories of objects. The performance of this improvement proves that incorporating spiking neurons with conventional convolutional layers can address the drawbacks of the performance of fully spiking models at a reasonable cost of energy consumption.

The other important strength of the Hybrid R-SNN is the greater recall in most object classes. This implies that the model will not be more bound to miss objects in a scene, but it will bring more false positives. This is a reasonable trade-off in terms of safety, particularly in systems that are used in autonomous driving, as missed detections are more dangerous than more detections. The recall oriented behavior is a design choice whereby the element of coverage is given more emphasis than exactness.

The effect of height-based analysis also indicates the efficiency of the Hybrid R-SNN in the hard case. Although the two models can work effectively with large and medium-sized objects, the Hybrid R-SNN has a higher retention of detection with small and distant objects. The temporal dynamics of spiking neurons can be used to explain this advantage, which enables weak visual signals to be accumulated over time and made detectable.

Regarding efficiency, Faster R-CNN has a lower inference latency, and Hybrid R-SNN has a higher power efficiency because it has an event-driven neuron activation.

The present Hybrid R-SNN implementation has only marginally extended inference time, but its energy-conscious implementation gives it an excellent basis to be optimized in the future.

All in all, Faster R-CNN still proves better in the cases when a high accuracy should be achieved, but Hybrid R-SNN is a viable advancement of traditional SNN-made detectors and a promising move in the direction of sustainable and energy-efficient object detection. As the Hybrid R-SNN is further refined in architecture and its training quality is enhanced, it can be assumed that the hybrid R-SNN will be able to achieve increased accuracy without sacrificing its energy-saving advantages.

Chapter 7

Conclusion

7.1 Summary of key findings

This research shows that implementing the dynamics of spiral neural within a region based convolutional neural network architecture is not only practical but also effective. The presented hybrid Faster R-SNN accomplishes the performance of strong detection on the KITTI dataset which achieves a mAP@0.5 OF 0.7395, mAP@0.7 of 0.5926 and COCO style mAP@[0.50:0.95] of 0.4771. Class wise evaluation demonstrates notably high Average Precision for primary vehicle classes for example- car, truck and van while performance decreases for rare and visually ambiguous classes such as person sitting. Furthermore, height based AP evaluation ensures that the model operates reliably on easy and moderate objects in line with expectations of challenges for small and heavily occluded objects, in particular pedestrians at extended ranges.

Apart from accuracy, the main contribution resides in efficiency. The model gets positive outcomes from temporal spike accumulation and event based computation which leads to high recall and decreased repetitive processing through the implementation of LIF based spiking neurons in the backbone of the Faster R-CNN. Efficiency metrics demonstrate a beneficial accuracy-time-energy trade off which emphasizes the applicability of the proposed hybrid Faster R-SNN model for edge devices especially for the applications of autonomous driving systems. In summary, the findings confirm that the hybrid SNN-DNN framework can maintain competitive detection accuracy while also delivering promising benefits in energy efficiency which makes them a potential candidate for low power vision systems in real time.

7.2 Contributions to the field

The hypothesis of the thesis is: An energy efficient hybrid Spiking Neural Network-based Faster R-CNN object detection. Rather than rearchitecturizing the whole architecture, Faster R-CNN key internal layers are implemented in terms of spike-based processing, but leave feature extraction, region proposal, and classification. The event-driven spiking neuron model consumes less power and unnecessary computation, so it can be used in edge and embedded systems. The trade-off between energy efficiency and detection accuracy is also presented by the study, which proves that reliable object detection can be implemented at lower computation cost and expands the use of spiking neural networks to full object detection pipelines.

In this study, we contribute to energy-efficient computer vision by addressing several key aspects. First, we create a hybrid spiking Faster R-CNN by converting selected internal layers to spike-based processing while keeping the overall detection framework intact. Second, we design an energy-efficient object detector utilizing event-driven spiking neurons which reduces computation and power consumption for edge devices. Third, we analyze the accuracy–energy trade-off, showing that our model achieves good detection performance at lower computational cost. Fourth, we demonstrate improved recall in safety-critical applications such as autonomous driving and surveillance. Finally, we extend the application of spiking neural networks from simple classification to full object detection, paving the way for scalable, energy-aware vision models on resource-constrained systems. In addition, the presented hybrid model attains higher detection accuracy than fully spiking models and also enhances the recognition of small and distant objects through temporal spike accumulation.

7.3 Recommendation for future work

This study has shown that a hybrid Spiking-Deep Neural Network could be used to achieve object detection with higher energy saving than other traditional deep learning models. The proposed system balances between the detection and lower computational cost by combining the spiking neural networks with deep neural architectures, which makes the system applicable to the low-power environment as well as on the edge-based systems. The comparison proves the future of neuromorphic-inspired systems in terms of efficient and sustainable visual perception systems.

To continue working on the model in the future, one can further explore more complex spiking neuron models and more sophisticated hybrid structures to further improve detection. More large-scale datasets like MS COCO or event-based vision datasets may be utilized to test generalization in real-world conditions. Besides, the implementation of the proposed model on the neuromorphic hardware systems and training optimization strategies may pave the way to better understanding the characteristics of real-time performance and energy consumption in practice.

Bibliography

[1] B. Zhou and J. Jiang, *Deep event-based object detection in autonomous driving: A survey*, https://arxiv.org/abs/2405.03995, Accessed: Jun. 14, 2025, 2024.

[2] S. Reddy, N. Pillay, and N. Singh, "Comparative evaluation of convolutional neural network object detection algorithms for vehicle detection," *Journal of Imaging*, vol. 10, no. 7, p. 162, 2024. DOI: 10.3390/jimaging10070162.

[3] A. Tavanaei, M. Ghodrati, S. R. Kheradpisheh, T. Masquelier, and A. Maida, "Deep learning in spiking neural networks," *Neural Networks*, vol. 111, pp. 47–63, 2019. DOI: 10.1016/j.neunet.2018.12.002.

[4] M. Jin, X. Wang, C. Guo, and S. Yang, "Research on target detection for autonomous driving based on ecs-spiking neural networks," *Scientific Reports*, vol. 15, no. 1, pp. 1–12, Apr. 2025. DOI: 10.1038/s41598-025-97913-4.

[5] E. Lielamurs et al., "A distributed time-of-flight sensor system for autonomous vehicles: Architecture, sensor fusion, and spiking neural network perception," *Electronics*, vol. 14, no. 7, p. 1375, Mar. 2025. DOI: 10.3390/electronics14071375.

[6] L. Fan, J. Yang, L. Wang, J. Zhang, X. Lian, and H. Shen, "Efficient spiking neural network for rgb–event fusion-based object detection," *Electronics*, vol. 14, no. 6, p. 1105, Mar. 2025. DOI: 10.3390/electronics14061105.

[7] G. Baris, B. Li, P. H. Chan, C. A. Avizzano, and V. Donzella, "Automotive dnn-based object detection in the presence of lens obstruction and video compression," *IEEE Access*, vol. 13, pp. 36 575–36 589, 2025. DOI: 10.1109/access.2025.3544773.

[8] D. Kang et al., *Real time scheduling framework for multi object detection via spiking neural networks*, https://doi.org/10.48550/arXiv.2501.18412, 2025.

[9] D.-H. Paek and S.-H. Kong, *Spikingrtnh: Spiking neural network for 4d radar object detection*, https://arxiv.org/abs/2502.00074, Accessed: Jun. 14, 2025, 2025.

[10] N. Saeedizadeh, M. Jafar, B. Khan, and S. Mohamed, "Cutting-edge deep learning methods for image-based object detection in autonomous driving: In-depth survey," *Expert Systems*, vol. 42, no. 4, Feb. 2025. DOI: 10.1111/exsy.70020.

[11] S. H. Ahmed, J. Finkbeiner, and E. Neftci, *Efficient event-based object detection: A hybrid neural network with spatial and temporal attention*, https://arxiv.org/abs/2403.10173, Accessed: Jun. 14, 2025, 2024.

[12] J. Qu, Z. Gao, T. Zhang, Y. Lu, H. Tang, and H. Qiao, "Spiking neural network for ultralow-latency and high-accurate object detection," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 36, no. 3, pp. 4934–4946, Mar. 2024. DOI: 10.1109/tnnls.2024.3372613.

[13] H. Zhang et al., "Automotive object detection via learning sparse events by spiking neurons," *IEEE Transactions on Cognitive and Developmental Systems*, pp. 1–15, Jan. 2024. DOI: 10.1109/tcds.2024.3410371.

[14] J. Courtois, P.-E. Novac, E. Lemaire, A. Pegatoquet, and B. Miramond, *Embedded event based object detection with spiking neural network*, https://arxiv.org/abs/2406.17617, Accessed: Jun. 14, 2025, 2024.

[15] T. Li et al., "An fpga accelerator design of spiking neural network for energy-efficient object detection," *IEEE Transactions on Consumer Electronics*, pp. 1–1, 2024. DOI: 10.1109/tce.2024.3502112.

[16] C. Iaboni and P. Abichandani, "Event-based spiking neural networks for object detection: A review of datasets, architectures, learning rules, and implementation," *IEEE Access*, vol. 12, pp. 180 532–180 596, 2024. DOI: 10.1109/access.2024.3479968.

[17] R.-J. Zhu, Z. Wang, L. Gilpin, and J. Eshraghian, *Autonomous driving with spiking neural networks*, https://proceedings.neurips.cc/paper_files/paper/2024/file/f7344147dbd1607deac3a7e5f33a23aa-Paper-Conference.pdf, Accessed: Jun. 14, 2025, 2024.

[18] X. Jin, M. Zhang, R. Yan, G. Pan, and D. Ma, "R-snn: Region-based spiking neural network for object detection," *IEEE Transactions on Cognitive and Developmental Systems*, vol. 16, no. 3, pp. 810–817, Sep. 2023. DOI: 10.1109/tcds.2023.3311634.

[19] Y. K. Wang, S. E. Wang, and P. H. Wu, "Spikeevent object detection for neuromorphic vision," *IEEE Access*, vol. 11, pp. 5215–5230, 2023. DOI: 10.1109/ACCESS.2023.3236800.

[20] S. A. Martinez, D. Ser, and P. Garcia-Bringas, *Efficient object detection in autonomous driving using spiking neural networks: Performance, energy consumption analysis, and insights into open-set object discovery*, https://arxiv.org/abs/2312.07466, Accessed: Jun. 14, 2025, 2023.

[21] N. Guzhva, V. Prun, R. Sadekov, V. Postnikov, and D. Sholomov, "Using 3d object detection dnn in an autonomous tram to predict the behaviour of vehicles in the road scene," in *2022 International Conference on Information and Navigation Systems (ICINS)*, 2022. DOI: 10.23919/ICINS51784.2022.9815388.

[22] S. Xiang et al., *Spiking siamfc++: Deep spiking neural network for object tracking*, https://arxiv.org/abs/2209.12010, 2022.

[23] S. Mohapatra, T. Mesquida, M. Hodaei, S. Yogamani, H. Gotzig, and P. Mader, *Spikili: A spiking simulation of lidar based real-time object detection for autonomous driving*, https://arxiv.org/abs/2206.02876, Accessed: Jun. 14, 2025, 2022.

[24] B. Chakraborty, X. She, and S. Mukhopadhyay, “A fully spiking hybrid neural network for energy-efficient object detection,” *IEEE Transactions on Image Processing*, vol. 30, pp. 9014–9029, 2021. DOI: 10.1109/tip.2021.3122092.

[25] F. Nezhadalinaei, L. Zhang, M. Mahdizadeh, and F. Jamshidi, “Motion object detection and tracking optimization in autonomous vehicles in specific range with optimized deep neural network,” in *2021 7th International Conference on Web Research (ICWR)*, 2021, pp. 53–63. DOI: 10.1109/ICWR51868.2021.9443120.

[26] K. Zhang, S. J. Wang, L. Ji, and C. Wang, “Dnn based camera and lidar fusion framework for 3d object recognition,” in *Journal of Physics: Conference Series*, vol. 1518, Apr. 2020, p. 012 044. DOI: 10.1088/1742-6596/1518/1/012044.

[27] S. Ren, K. He, R. Girshick, and J. Sun, “Faster R-CNN: Towards real-time object detection with region proposal networks,” *arXiv preprint arXiv:1506.01497*, 2015. [Online]. Available: https://arxiv.org/abs/1506.01497.

[28] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, “Vision meets robotics: The kitti dataset,” *International Journal of Robotics Research (IJRR)*, 2013.

[29] O. Ghai, “Investigating and comparing yolo models for roadside object detection,” *NHSJS*, 2026, Accessed: 2026-02-03. [Online]. Available: https://nhsjs.com/2026/investigating-and-comparing-yolo-models-for-roadside-object-detection/.

[30] P. Abillama et al., “One-hot multi-level leaky integrate-and-fire spiking neural networks for enhanced accuracy-latency tradeoff,” *IEEE Access*, vol. 13, pp. 38 163–38 180, 2025. DOI: 10.1109/access.2025.3546508. [Online]. Available: https://doi.org/10.1109/access.2025.3546508.

[31] ML Katas, *Spiking neuron with leaky integrate-and-fire — kata 84*, https://mlkatas.com/kata/84, Accessed: 2026-01-28, 2026.

[32] J. Wu, Y. Wang, Z. Li, L. Lu, and Q. Li, “A review of computing with spiking neural networks,” *Computers, Materials & Continua*, vol. 78, no. 3, pp. 2909–2939, 2024. DOI: 10.32604/cmc.2024.047240. [Online]. Available: https://doi.org/10.32604/cmc.2024.047240.

[33] D. K. Alqahtani, A. Cheema, and A. N. Toosi, “Benchmarking deep learning models for object detection on edge computing devices,” *arXiv preprint arXiv:2409.16808*, 2024. [Online]. Available: https://arxiv.org/abs/2409.16808.

[34] U. Authors, “Energy-efficient deep learning models: Metrics, benchmarks, and applications,” *Journal of Real-Time Image Processing*, 2025. DOI: 10.1007/s11554-025-01703-0. [Online]. Available: https://doi.org/10.1007/s11554-025-01703-0.

[35] E. Cai, D.-C. Juan, D. Stamoulis, and D. Marculescu, “Neuralpower: Predict and deploy energy-efficient convolutional neural networks,” *arXiv preprint arXiv:1710.05420*, 2017. [Online]. Available: https://arxiv.org/abs/1710.05420.