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

Asrar Azim
Computer Science and Engineering
Brac University
Dhaka, Bangladesh
asrar.azim@g.bracu.ac.bd

Md. Arif Uz Zaman Sakib
Computer Science and Engineering
Brac University
Dhaka, Bangladesh
arif.uzzaman.sakib@g.bracu.ac.bd

Md. Rafid Jamal
Computer Science and Engineering
Brac University
Dhaka, Bangladesh
md.rafid.jamal@g.bracu.ac.bd

Nabiha Binte Yousuf
Computer Science and Engineering
Brac University
Dhaka, Bangladesh
nabiha.binte.yousuf@g.bracu.ac.bd

Shihana Sultana Jhilik
Computer Science and Engineering
Brac University
Dhaka, Bangladesh
shihana.sultana.jhilik@g.bracu.ac.bd

***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.**

***Index Terms*—Spiking Neural Network (SNN), Faster R-CNN, Object detection, Energy efficiency , KITTI dataset, Edge computing, Autonomous driving.**

I. Introduction

A. Background

Autonomous vehicles are smart systems that are able to make real time driving decisions based on sensor data via cameras, LiDAR and radar. Their purpose is to create better road safety by lowering human error and supporting new drivers; also creating mobility to people who cannot drive. Autonomous systems will reduce accidents and enhance efficiency in traffic by predicting traffic conditions and acting before it happens [1].

Although such benefits are present, fully autonomous cars are in the development stage. The existing systems have some issues connected to reliability, ethical choice, and energy efficiency. Even though the majority of autonomous vehicles are electric, they make their heavy use of high-performance sensors and models that need high performance and computations quite concerns in terms of power consumption. Autonomous vehicles can be more environmentally sustainable than traditional ones in case they are powered by renewable energy sources and used in a more energy-friendly manner through computers.

B. Motivation

Object detection is an essential activity in computer vision and it is vital in applications like autonomous driving. Deep Neural Networks (DNNs), such as YOLO, SSD, and Faster R-CNN, are highly accurate since they are capable of learning rich features out of data and can manage variations in scale, lighting, and object orientation [2]. Nonetheless, such models are very demanding in terms of computational power and energy, and thus, are not applicable to resource constrained and edge devices.

A Spiking Neural Networks (SNNs) is a more energy efficient alternative in that it represents a biological neural communication using spatially localized spikes. SNNs consume very little power unlike DNNs since they only process information when there are changes. The main driving force behind this is to use the energy efficiency of SNNs and preserve the accuracy of DNNs in object detection of autonomous systems.

C. Problem Statement

Autonomous vehicles depend on the continuous perception of the environment to execute functions like object detection, lane detection, and predicting the traffic. Although DNN-based models are highly accurate, they continuously update over time, irrespective of the change in the environment and use high energy which restricts their use in systems with low power.

However, SNNs are event-driven and also use much less energy but have lower accuracy on difficult vision tasks. The answer lies in hybrid SNN-DNN architectures which use the accuracy of DNNs and the energy efficiency of SNNs. The main problem that has been dealt with in this paper is how to design a hybrid object detection model that can be used in real-time autonomous and edge-based applications, is fast, accurate, and energy-efficient at the same time.

D. Objectives

The primary task of this study is to create and test a hybrid Spiking-Deep Neural Network (SNN-DNN) based Faster R-CNN object detector that is precise and power efficient. The specific objectives are:

- To create a hybrid SNN-DNN Faster R-CNN object detector to work in real-time.
- To attain a high detection rate and low inference time.
- To test the model on a conventional object detection dataset.
- To examine performance regarding accuracy, latency, energy requiring consumption, and complexity of the model.
- To validate the relevance of hybrid SNN-DNN models in the low-power and edge devices.

E. Methodology in brief

The proposed methodology has four key steps:

- Dataset Preparation: The KITTI dataset is selected because it is relevant to autonomous driving situations. The data will be split into training and testing sets and data augmentation strategies like scaling, flipping, and normalization will be used to enhance generalization.
- Model Implementation: Two object detection models are deployed, standard Faster R-CNN with DNN and hybrid Faster R-CNN with SNN backbone.
- Training and Evaluation: The two models are both trained and tested on the same data splits in order to make a fair comparison. Mean Average Precision (mAP), Intersection over Union (IoU) and inference latency are used to determine performance.
- Comparative Analysis: The models are contrasted based on accuracy of detection, inference speed, energy usage, and complexity of computation to assess the performance-efficiency ratio and resource-constrained autonomous systems appropriateness.

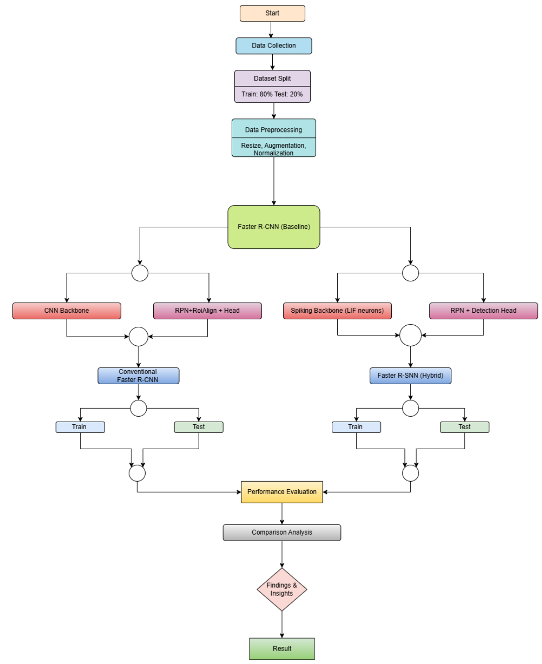

Fig. 1. Work Plan

F. Research Contribution

The thesis gives a contribution to the energy efficient computer vision by presenting a hybrid Spiking-Deep Neural Network (SNN-DNN) object detection model. The proposed solution does not complete the replacement of the conventional deep learning models, but it involves the spiking neural computation into the chosen internal units of the Faster R-CNN architecture, which reduces the energy use and maintains the level of detection accuracy.

The key findings of the research are as follows:

- Hybrid spiking Faster R-CNN architecture: Shows that it is possible to convert Faster R-CNN to a hybrid, SNN-based detector without changing its overall detection pipeline, and to substitute with spike-based processing prescribed individual network layers.
- Energy efficient object detection: The energy-efficient model using event-driven spiking neurons to minimize unnecessary computation and energy use, therefore, can be applied to edge and low-power devices.
- Better accuracy compared to fully spiking models: A combination of convolutional feature extraction and spiking computation is both more accurate than fully SNN-based detectors.
- Better detection of small and distant objects: Temporal accumulation of spiking neurons enhances weak visual signal detection leading to a higher performance of small and far objects than the traditional Faster R-CNN.
- Analysis Accuracy-energy trade-off: A specific analysis which gives an in-depth analysis of detection and energy efficiency and shows competitive accuracy at reduced computational cost.

- Better recall of safety critical applications: Demonstrates better recall on object classes lessening miss-detections in safety critical applications, needed by autonomous driving and surveillance systems.
- Generalization of SNNs to larger detection systems: Extends the use of SNNs to complex object detection systems to novel studies into scalable and energy efficient vision systems.

G. Scopes and challenges

This paper aims at designing, implementing as well as evaluating a hybrid SNN-DNN object detector model to use in real time autonomous as well as edge based applications. The proposed hybrid architecture based on faster R-CNN is designed to maximize its accuracy and latency and use less energy, and therefore it is applicable in resource-bound systems. The standard object detection metrics, speed of inference, use of energy and complexity of the model are used to assess model performance using benchmark datasets that are pertinent to autonomous driving.

Although it has its benefits, there are a number of challenges that the development of the hybrid SNN-DNN object detectors have:

- The complexity of training: making spike-based learning is challenging in terms of optimization relative to traditional training of DNNs.
- Performance-efficiency trade-off: To attain high levels of accuracy with the lowest level of latency and power consumption, it is important to design it carefully.
- Poor software and hardware support: SNN systems are still immature and the hardware that is able to support neuro-morphology is still not readily available.

These issues need to be solved in order to come up with viable, rapid, and efficient energy-saving object detection systems applicable in real-world autonomous and edge computing use.

II. Related Work

ECSLIF-YOLO combines AI frameworks of ECS-LIF spiking neurons and the YOLO to enhance object recognition in autonomous driving [3]. 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 [4], 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 [5] 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 [6]. 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 [7], 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 [8]. 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 [9]. 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 [10]. 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 [11] 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 [12]. 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 [13]. 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 [14], 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 [15]. 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 [16]. 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 [17] 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 [18]. 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 [19]. 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.

[20] 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 [21]. 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 [22]. 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 [23], 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 [24]. 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 [25]. 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 [26].

Recent developments in spiking neural networks (SNNS) show strong potential for energy efficient object detection in autonomous driving systems while preserving competitive accuracy. Multiple SNN and hybrid SNN-ANN models have attained high mAP with substantial reductions in power consumption utilizing neuromorphic hardware and multimodal sensors. Energy efficiency enhancements of up to several hundred times over ANN baselines have been observed in recent studies. Practical implementations on FPGA and neuromorphic chips further verify real time and low power applicability. Despite this, constraints persist in challenging condition robustness, small object detection, accuracy and implementation complexity.

III. Proposed Method

This study presents a hybrid Faster R-SNN for energy efficient object detection by integrating spiking neuron dynamics into the conventional Faster R-CNN architecture. The objective is to introduce temporal and event-driven computation while maintaining the accuracy of the localization and performance of the detection of conventional Faster R-CNN.

A. Hybrid Faster R-SNN Architecture

Spiking Neural Network (SNN) processing is integrated into the feature extraction stage to decrease the computational cost and power usage. The CNN backbone is unaltered and conventional feature maps which are then handled by Leaky Integrate and Fire (LIF) neurons to produce spike based representations. The Region Proposal Network (RPN) and detection heads perform operations in the analog domain that permits the integration of spiking computation without altering the overall detection pipeline. The presented Faster R-SNN model's complete pipeline along with ANN modules, spiking modules and spike to spike transition is displayed in figure 2.

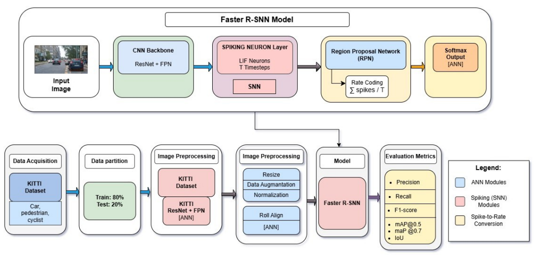

Fig. 2. R-SNN Architecture Diagram

B. LIF Neuron Model

Spiking behaviour is represented utilizing Leaky Integrate and Fire neurons. 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, 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 [27]. A spike is produced when the membrane potential exceeds a threshold, then the potential is reset. The figure 3 illustrates the manner in which the neuron receives the input signals, generates spikes and restores its condition in the course of time.

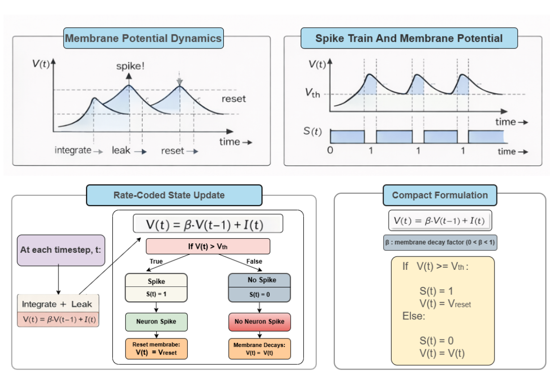

Fig. 3. Internal Dynamics of LIF neuron illustrating membrane integration, spike generation and reset mechanism over time

Spike activity is represented utilizing rate coding where spike counts represent activation magnitude as time progresses. This facilitates temporal information implementation while remaining compatible with the layers of analog.

C. Data Preprocessing

The KITTI dataset is preprocessed through organizing image-annotation pairs, mapping object classes to numerical labels and eliminating invalid samples. All valid annotations are transformed into structured numerical formats. The dataset is divided into 80% training and 20% testing. Images are transformed to RGB format, augmented as needed and batched utilizing a custom data loader.

D. Implementation of selected design

1) Backbone and spiking layer: The backbone mainly relies on a ResNet with Feature Pyramid Network (FPN) that generates multiscale feature maps for detecting objects of different sizes. A spiking neuron layer is positioned after the backbone that substitutes conventional activations with LIF neurons. Feature activations are handled over T timesteps which transform static features into temporal spike trains without developing recurrent connections.

2) Spike to rate conversion: Spike outputs are computed as the mean over time to interface spiking and non-spiking components:

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

This rate coded representation maintains temporal information at the same time facilitating convolutional and regression operations [28].

3) RPN and detection head: The RPN handles rate coded features to produce object proposals which is followed by RoIAlign and detection heads for classification and bounding box refinement. These phases remain fully analog to maintain stable optimization and accurate localization.

IV. Results and Discussion

A. Experimental Setup

All experiments were conducted on the KITTI object detection dataset following the standard train–test split. The evaluation was performed using commonly adopted object detection metrics, including mean Average Precision (mAP), Precision, Recall, and F1-score. The RSNN and RCNN models were trained under identical conditions to ensure a fair comparison. The primary objective of this evaluation is to assess the effectiveness of the proposed RSNN framework in terms of detection accuracy and learning behavior, while benchmarking it against a conventional Faster R-CNN baseline.

B. Performance of the Proposed Model

This subsection presents a detailed performance analysis of the proposed hybrid object detection model.

1) Overall Detection Performance: The hybrid model achieves reasonably strong detection performance across all evaluated metrics, demonstrating its ability to effectively localize and classify objects in complex traffic scenes.

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

This result confirms that incorporating spiking neuron dynamics into a region-based detection framework does not compromise detection quality, while enabling temporal and event-driven feature processing.

2) Learning Behavior and Convergence: The training dynamics of the hybrid model are illustrated in Fig. 4, which shows the evolution of training and validation loss and accuracy over epochs. The model exhibits stable convergence, with a consistent decrease in loss and a gradual improvement in detection accuracy.

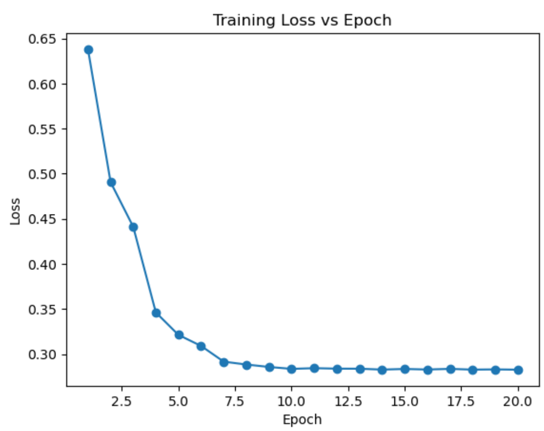

Fig. 4. Training Loss Vs Epoch (Hybrid Model)

This behavior indicates that the surrogate gradient-based learning strategy is effective in optimizing the spiking network and preventing instability typically associated with non-differentiable spike functions.

3) Class-wise Performance and Confusion Analysis: To further analyze the classification behavior of the hybrid model, a confusion matrix is presented in Fig. 5. The matrix reveals that the model achieves high recognition rates for dominant object classes such as car, truck, and cyclist, while relatively lower performance is observed for visually ambiguous or underrepresented classes such as person sitting.

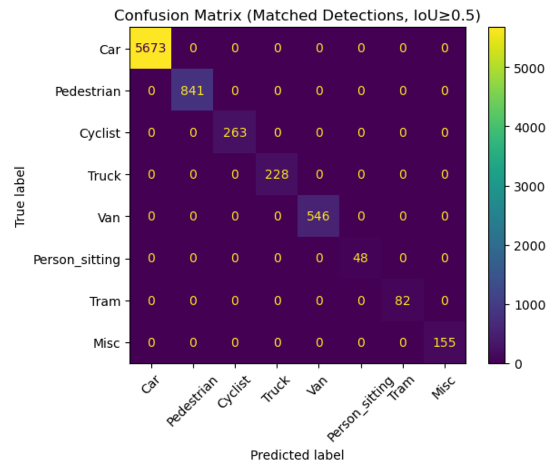

Fig. 5. Confusion Matrix

This outcome aligns with the class imbalance present in real-world autonomous driving datasets and highlights the robustness of the model in learning discriminative spatial–temporal features for common object categories.

4) Qualitative Detection Results: Qualitative detection samples are illustrated in Fig. 5, showing predicted bounding boxes and class labels overlaid on real driving scenes. The hybrid model demonstrates accurate localization and consistent confidence levels across different object scales and environmental conditions.

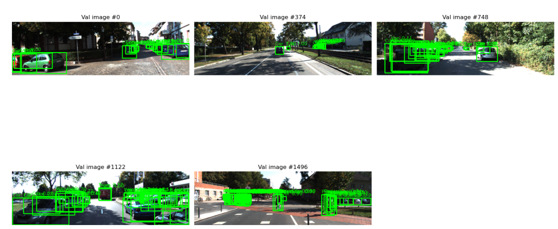

Fig. 6. Sample Detection

These results visually confirm the quantitative findings and validate the practical effectiveness of the proposed framework.

C. Comparison with RCNN Baseline

To benchmark the proposed RSNN model, a comparative evaluation was conducted against a conventional Faster R-CNN baseline trained under the same experimental conditions.

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

The Faster R-CNN achieves a higher overall mAP and slightly better precision, reflecting its strong spatial feature extraction capability. However, the RSNN demonstrates competitive detection performance while introducing temporal processing and event-driven computation. Although the RSNN exhibits marginally lower raw accuracy, it achieves comparable recall and F1-score, indicating a robust ability to detect relevant objects without excessive false negatives.

Importantly, the RSNN offers additional advantages that are not captured solely by accuracy metrics. The spiking architecture enables sparse neuron activation and temporal accumulation of visual information, making it inherently more suitable for energy-efficient and real-time edge deployment scenarios.

D. 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 [29]. 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 [30], [31]. 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.

E. Discussion

The experimental results highlight a fundamental trade-off between conventional deep learning-based object detection and spiking neural approaches. While Faster R-CNN remains superior in terms of peak detection accuracy, the proposed hybrid model achieves competitive performance with significantly reduced computational activity.

The model also demonstrates improved robustness in scenarios involving small or partially occluded objects, which can be attributed to the temporal integration capability of spiking neurons. This property allows weak visual signals to accumulate over time, enhancing detectability in challenging conditions.

Overall, these findings suggest that the proposed Hybrid R-SNN framework represents a promising step toward sustainable and energy-efficient object detection systems. With further architectural optimization and hardware-level implementation, RSNN-based detectors have strong potential to bridge the performance gap with conventional deep neural networks while offering substantial gains in energy efficiency and scalability.

V. Conclusion

A. Summary of key bindings

This work demonstrates that integrating spiral spiking neural dynamics into a region-based convolutional neural network is both feasible and effective. The proposed hybrid Faster R-SNN achieves strong detection performance on the KITTI dataset, with 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 shows high accuracy for dominant vehicle classes such as car, truck, and van, while reduced performance is observed for rare or visually ambiguous classes. Height-based evaluation confirms reliable detection for easy and moderate objects, with expected degradation for small and heavily occluded targets.

Beyond accuracy, the primary contribution lies in efficiency. The use of LIF-based spiking neurons enables event-driven computation and temporal spike accumulation, reducing redundant processing. Experimental results highlight a favorable accuracy–latency–energy trade-off, demonstrating the suitability of the hybrid Faster R-SNN for energy-constrained edge devices, particularly in autonomous driving applications.

B. Contributions to the field

This study proposes an energy-efficient hybrid Faster R-CNN by selectively integrating spike-based processing into key internal layers while preserving the overall detection framework. The contributions include:

- The design of a hybrid SNN–DNN object detector.
- Reduced computational and power consumption through event-driven spiking neurons.
- Analysis of the accuracy–energy trade-off.
- Improved recall for safety-critical applications.
- Extension of spiking neural networks from classification to full object detection pipelines.

The proposed model outperforms fully spiking approaches in detection accuracy while enhancing small and distant object recognition through temporal spike accumulation.

C. Recommendations for future work

Future research may explore more advanced spiking neuron models and hybrid architectures to further improve detection performance. Evaluation on larger-scale datasets such as MS COCO and event-based vision benchmarks would help assess generalization. Additionally, deployment on neuromorphic hardware and investigation of training optimization strategies could provide deeper insights into real-time performance and practical energy efficiency.

References

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

[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.

[3] 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.

[4] 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.

[5] 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.

[6] 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.

[7] 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.

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

[9] 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.

[10] 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, 2024, accessed: Jun. 14, 2025.

[11] 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.

[12] 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.

[13] 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, 2024, accessed: Jun. 14, 2025.

[14] 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.

[15] 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.

[16] 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, 2024, accessed: Jun. 14, 2025.

[17] 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.

[18] Y. K. Wang, S. E. Wang, and P. H. Wu, "Spikeevent object detection for neuromorphic vision," *IEEE Access*, vol. 11, pp. 5215–5230, 2023.

[19] 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, 2023, accessed: Jun. 14, 2025.
[20] 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.
[21] S. Xiang *et al.*, “Spiking siamfc++: Deep spiking neural network for object tracking,” https://arxiv.org/abs/2209.12010, 2022.
[22] 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, 2022, accessed: Jun. 14, 2025.
[23] 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.
[24] 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.
[25] 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. 012044.
[26] 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
[27] ML Katas, “Spiking neuron with leaky integrate-and-fire — kata 84,” https://mlkatas.com/kata/84, 2026, accessed: 2026-01-28.
[28] 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. [Online]. Available: https://doi.org/10.32604/cmc.2024.047240
[29] 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
[30] U. Authors, “Energy-efficient deep learning models: Metrics, benchmarks, and applications,” *Journal of Real-Time Image Processing*, 2025. [Online]. Available: https://doi.org/10.1007/s11554-025-01703-0
[31] 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