MLRDv2: A Dataset for Improving Micromobility Safety via Attention-Integrated Compact CNN Models

Chinmaya Kaundanya[1*], Paulo Cesar[2], Barry Cronin[2], Andrew Fleury[2], Mingming Liu[1], Suzanne Little[1]

[1*]Research Ireland Insight Centre for Data Analytics, Dublin City University, Dublin, D09 V209, Ireland.
[2]Luna Systems, Dublin, D11 KXN4 Ireland.

*Corresponding author(s). E-mail(s): chinmaya.kaundanya3@mail.dcu.ie;
Contributing authors: paulo.cesar@luna.systems;
barry.cronin@luna.systems; andrew.fleury@luna.systems;
mingming.liu@dcu.ie; suzanne.little@dcu.ie;

Abstract

Micromobility offers a sustainable alternative to traditional transportation but lacks clear safety regulations. Existing sensor-based solutions for micromobility safety are often imprecise, expensive, or resource-intensive, making them unsuitable for constrained environments. AI-based lane detection techniques hold potential, but most rely on image segmentation, which is computationally demanding. A significant challenge is the absence of dedicated image datasets for micromobility, as current datasets primarily focus on autonomous driving and do not capture the unique perspectives of micromobility vehicles. To address this, we introduce the Micromobility Lane Recognition Dataset (MLRD), which enables real-time lane identification to regulate rider behavior. Using MLRD, we evaluate the effect of channel and spatial attention mechanisms on compact convolutional neural networks (CNNs). Our findings show that combining channel and spatial attention improves CNN performance by enabling better focus on important features. MobileNet V2, with integrated attention mechanisms, achieved the highest precision and F1 scores, while MobileNet V3 maintained strong performance with fewer parameters. To meet the growing demand for micromobility, we also present MLRDv2, an improved dataset featuring more diverse scenarios. Testing on MobileNet V2 and V3 Large models showed a 4% performance boost compared to results from MLRD V1, demonstrating the dataset's effectiveness.

Keywords: Micromobility, Lane Recognition, MobileNet, Attention Mechanisms

1 Introduction

Micromobility offers a transformative and innovative solution to reduce the reliance on private vehicles for short-distance travel by utilizing personal mobility devices such as e-bikes and e-scooters. This approach provides a sustainable, flexible, and cost-efficient alternative to traditional carbon-powered transportation. As micromobility gains increasing popularity globally, the need to develop and implement effective regulations for the safe use of e-scooters and e-bikes has become increasingly important.

According to a recent report by the Insurance Institute for Highway Safety (IIHS) [1], 60% of e-scooter accidents occur on sidewalks, prompting cities to mandate operators to have robust safety technologies as a prerequisite for micromobility licensing. Ensuring the safe and widespread use of micromobility solutions requires not only regulatory frameworks but also advanced technological supports, such as highly accurate, real-time lane recognition systems. Existing GPS and sensor technologies struggle to provide the necessary precision or adapt to the diverse road structures encountered by micromobility vehicles [2]. While LASER and LiDAR sensors offer potentially more accurate solutions, they are computationally expensive and challenging to deploy in the constrained environments where micromobility vehicles operate [3]. Computer Vision based techniques using image inputs present a promising alternative but currently perform optimally only in controlled environments (e.g., high light exposure, stable camera perspective) and require high-spec computational and memory resources.

To address the gap between current challenges and the needs of micromobility, we focus on the state-of-the-art deep-learning-based lane recognition techniques. Most existing techniques for lane recognition in autonomous driving rely on semantic segmentation [4], which is not well-suited for the constrained environments of micromobility. Segmentation models are typically large, requiring substantial memory and computational power for fast inference. However, micromobility vehicles, being more affordable and limited in physical space and power capacity, have significantly lower computational resources, making such models impractical for deployment.

Considering the resource limitations and the specific challenges of micromobility safety, we propose a lane recognition strategy that utilizes channel and spatial attention mechanisms in Convolutional Neural Networks (CNNs). This approach is designed to accurately detect the lane in which a micromobility rider is traveling in real-time, enabling the system to provide timely alerts as necessary.

Attention mechanisms in CNN enable the model to focus on the most relevant input features in an image, playing a crucial role in enhancing both performance and efficiency [5, 6]. The two commonly used types of attention mechanisms, spatial and channel attention, capture pixel-level pairwise relationships and channel dependencies, respectively [7]. The goal of incorporating these attention algorithms is to optimize the performance of compact CNN models, particularly in resource-constrained environments, by adding a minimal number of trainable parameters while maintaining computational efficiency.

One of the key challenges in developing lane recognition solutions for micromobility is the absence of specialized, labelled datasets containing images from a first-person or rider's perspective. Previous research on urban footpaths used crowd-sourcing to create a dataset specifically for urban mobility analysis, emphasizing the shortcomings of existing large datasets. Existing datasets primarily consist of images of footpaths and bike lanes captured from cameras mounted on cars that are not fully representative of the micromobility perspective [8].

To overcome these challenges, we developed the Micromobility Lane Recognition Dataset (MLRD), a novel multi-label image classification dataset specifically designed for micromobility safety applications, particularly for real-time lane recognition from the rider's first-person perspective. This dataset provides crucial information about whether the rider is traveling on a road, bike lane, or sidewalk.

In this work, we expand on our initial efforts by introducing Micromobility Lane Recognition Dataset Version 2 (MLRDv2), an enhanced version of the original dataset. This improvement was achieved by curating and filtering MLRD Version 1, while also broadening its scope with additional frames from cities in Europe and the UK. The expanded dataset consists of frames from more complex and realistic scenes, such as cloudy, rainy weather, and nighttime conditions, making it more comprehensive for training and testing lane recognition image classification models. Details of the proposed dataset can be found in Section 7. As shown in Fig. 1, while the overall development pipeline includes model optimization and deployment on resourceconstrained platforms, this study focuses specifically on the dataset and model design stages.

This research serves as an initial investigation into the question: *What impact does adding computationally cheap operations such as channel and spatial attention have on the performance of compact CNN image classification models in constrained environments with low-resolution input images?* Given these limitations, accurate lane classification for micromobility requires the model to effectively capture both channel-wise dependencies and spatial relationships within the image feature maps. Based on this understanding, we selected Squeeze-and-Excitation [9] to integrate channel based attention, and Coordinate Attention [10], which provides a more comprehensive approach, incorporating both channel and spatial attention.

In this study, we conducted a series of experiments to evaluate the effect of integrating attention mechanisms into compact CNN models for image classification. Specifically, the MobileNetV3 [11] model, integrated with with both channel and spatial attention (Coordinate Attention), exhibited remarkably stable performance metrics, achieving results nearly on par with the baseline model despite having fewer parameters. The MobileNetV2 [12] variant with the channel-based attention (Squeeze-and-Excitation) showed a modest improvement in overall precision, while MobileNetV2 integrated with the Coordinate Attention demonstrated a notable increase in precision and a slight improvement in overall performance.

However, the MobileNetV2 model with the Coordinate Attention exhibited a noticeable decline in the recall metric for the 'road' class, and the standard

MobileNetV3 model with Squeeze-and-Excitation attention showed a deterioration in overall performance. This pattern indicates that while attention mechanisms can help reduce false positives, they may also introduce the risk of overfitting, negatively affecting the models' overall performance. These results suggest that simply integrating channel and spatial attention mechanisms into compact base models does not consistently enhance performance. This underscores the need for more extensive research, particularly across diverse use cases, to better understand and optimize the integration of attention mechanisms in compact CNN architectures.

The contributions of this paper are as follows:

1. Proposing a lightweight lane recognition solution for micromobility vehicle and rider safety using compact multi-label image classification model that can identify road lane types in real-time when deployed on low-spec microcontrollers.
2. Addressing the limitations of existing image datasets for micromobility safety applications, this work highlights their target use-cases and underscores the lack of image classification datasets specifically dedicated to lane recognition for micromobility vehicle and rider safety.
3. Evaluating compact MobileNet models, both with and without channel and spatial attention on Micromobility Lane Recognition Dataset Version 1 (MLRD).
4. Introducing Micromobility Lane Recognition Dataset Version 2 (MLRDv2), a multi-label image classification dataset for real-time lane recognition for micromobility vehicle and rider safety, featuring more curated frames and increased variety compared to the previous version.

The paper is organized as follows: Section 2 provides a review of existing datasets focused on micromobility safety and lane recognition methods using attention mechanisms. Section 3 outlines the technical specifications of the models and algorithms used. In Section 4, we present the details of the Micromobility Lane Recognition Dataset (MLRD). The experimental methodology and implementation specifics are described in Section 5. Section 6 discusses the experimental results obtained on MLRD. We introduce and present experimental results on the new extended version of MLRD, referred to as MLRDv2, in Section 7. Finally, Section 8 offers conclusions and explores potential future work. Our code and dataset are publicly accessible at: https://github.com/Luna-Scooters/Compact-Attention-based-CNNs-on-MLRD

2 Related work on Micromobility safety

2.1 Datasets

Image datasets for micromobility safety primarily focus on capturing real-time scenes encountered by users, often by mounting cameras on e-scooters or e-bikes. Datasets such as PolyMMV [13] emphasize detecting objects like bicycles, e-scooters, and skateboards, addressing real-world challenges such as occlusions and motion blur. In contrast, the ScooterDet [14, 15] dataset is designed specifically to detect other e-scooter riders and pedestrians. The primary goal of both datasets is to detect moving objects in real-time and provide timely warnings to users to prevent potential collisions. Meanwhile, the Outdoor Hazard dataset [16] and Road Pothole Detection datasets [17, 18] are focused on detecting stationary and hazardous objects, including potholes, bumps, fences, walls, and traffic cones. These datasets are crucial for enhancing the safety of micromobility riders and vehicles.

In addition to visual datasets, there has been work in micromobility safety applications using multi-sensor data. For instance, the Dataset for multimodal transport analytics of smartphone users – Collecty [19] offers a detailed breakdown of raw data based on users, transport modes, and multimodal routes. During data collection, sensor data from accelerometers, magnetometers, and gyroscopes embedded in mobile devices were recorded. This dataset provides valuable insights into e-scooter usage patterns and behaviours that may contribute to accidents or unsafe conditions.

Finally there are private datasets like the Real-Time Vibration Sensor dataset [20] that collect vibration data using six-axis sensors mounted on e-scooters during driving sessions. This dataset is employed to classify driving states—safe, slightly anxious, and very anxious—using deep learning models. All these datasets provide valuable resources for developing AI solutions aimed at enhancing micromobility vehicle and rider safety. However, there remains a significant gap in the availability of datasets specifically dedicated to lane recognition in the context of micromobility safety.

2.1.1 Lane Recognition for Micromobility

Previous studies on lane recognition have begun utilizing attention mechanisms to enhance the accuracy of lane detection, segmentation, and classification. However, these approaches primarily focus on footage captured from cars and aim to detect the type and location (relative or absolute) of lanes in conventional road driving scenarios. Here, we specifically review methodologies that have incorporated attention mechanisms for lane recognition, and based on the strong outcomes observed in traditional automotive scenarios, we evaluate their suitability for low-resource micromobility applications.

Zhang et al. [21] introduced a real-time lane recognition system that leverages the Convolutional Block Attention Module (CBAM) [22], which applies both channel and spatial attention mechanisms. Their Convolutional Neural Network (CNN) architecture consists of an encoder for extracting lane-specific features, a binary decoder for lane classification, and an additional decoder to predict feature maps representing individual lane instances. The integration of CBAM allows the encoder to capture detailed information from the targeted regions more effectively. This approach creates a synergy between features extracted by the convolutional layers and those obtained through the attention mechanism, enhancing the model's ability to capture contextual information. This contextual knowledge is then combined with upsampled features in the decoders to recover any lost details. The binary decoder classifies pixels as either lane or non-lane, while the second decoder distinguishes between distinct lane instances. The system was tested on the TuSimple [23] and Caltech Lanes [24] datasets to evaluate its performance.

Building on the use of attention mechanisms in encoder-decoder-based image segmentation models, Li et al. [25] introduced the Lane-DeepLab model, aimed at improving high-definition map generation for autonomous driving. Their architecture incorporates a novel attention module within the Atrous Spatial Pyramid Pooling

(ASPP) module of the encoder, which enhances feature extraction, along with a Semantic Embedding Branch (SEB) that fuses high- and low-level semantic information for richer feature acquisition. By combining attention mechanisms with contextual semantics, the system effectively integrates relevant information to improve the accuracy of lane line detection. This approach enables the model to adapt to and accurately interpret complex and dynamically changing road conditions. The Lane-DeepLab model's performance was evaluated on the TuSimple [23] and CULane [26] datasets, demonstrating its effectiveness.

Focusing on spatial features, particularly the positions of various road segments, a robust lane detection method was developed using a novel self-attention module called Expanded Self Attention (ESA), optimized specifically for lane detection [27]. This method improves segmentation-based lane detection by extracting global contextual information. The ESA module is divided into Horizontal (HESA) and Vertical (VESA) components, which predict occluded lane positions by assessing lane confidence in both directions. Designed to address challenges such as occlusion and difficult lighting conditions, this approach was evaluated on three popular datasets: TuSimple [23], CULane [26], and BDD100K [28].

Considering the memory constraints of low-spec platforms, Yao et al. [29] proposed an efficient lane detection method using a lightweight, attention-based deep neural network. The architecture features two branches: the Global Context Embedding (GCE) branch, which captures overall lane information, and the Explicit Boundary Regression (EBR) branch, which integrates a Spatial Attention Mechanism (SAM) for precise boundary detection. Additionally, the network incorporates a Channel Attention Mechanism (CAM) to prioritize channels containing target objects. Notably, the model achieved a high processing speed of 259 frames per second (FPS) on an NVIDIA GTX 2070 GPU with an input image resolution of 640x360. Demonstrating efficiency, the model required only 1.57M parameters and was tested on the CULane dataset [26]. The approach was thoroughly evaluated on both the TuSimple [23] and CULane [26] datasets, resulting in the accuracy of 95.11% and 68.8% respectively.

Finally, while high-performing models can be compressed or pruned for deployment, recent techniques such as self-distillation [30] and channel pruning [31]—although designed for lightweight CNN-based lane detection [32]–often still produce models that are too large for stringent, low-spec, resource-constrained edge platforms. While attention mechanisms improve the performance of standard models with only a minor increase in parameters, leading to a slight rise in memory usage, existing research has not thoroughly explored their impact on compact CNN models. This is especially important for models operating under strict memory and inference speed constraints, as seen in low-spec edge platforms. Overcoming this gap is necessary to maximize the performance of the compact CNN models with the addition of minimal amount of trainable parameters.

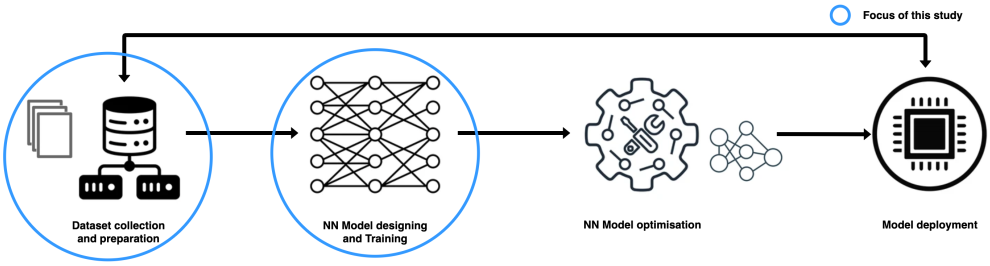


Fig. 1: A schematic representation of the workflow for developing and deploying a neural network model, illustrating the stages of data collection, model training, optimization, and final deployment on a target platform [36].

3 Models for Lane Recognition

3.1 MobileNets

MobileNets, a class of models designed for mobile and embedded vision tasks, have transformed image classification by striking a balance between model accuracy and computational efficiency [33]. Their architecture utilizes depthwise separable convolutions, which drastically lower the amount of computations while maintaining efficient feature extraction. With MobileNetV2 [12], the introduction of inverted residuals and linear bottlenecks further streamlined the CNNs, improving the flow of information and gradients during training. This advancement highlights the potential of lightweight, high-performance CNN models in resource-limited environments.

MobileNetV3 [11] introduced further refinements by leveraging automated Neural Architecture Search (NAS) algorithms [34], resulting in a more hardware-optimized design. This version, known for its enhanced speed and efficiency, is well-suited for real-time image classification tasks and facilitates the integration of AI in mobile and embedded systems. Additionally, the use of the Hard Swish activation function, along with an integrated channel attention mechanism, further improved performance over MobileNetV2 by boosting both efficiency and accuracy. A key feature of MobileNets is its architectural versatility, which allows for the seamless incorporation of individual algorithms, such as attention mechanisms, into the base model, increasing its functionality and adaptability across various computational settings [35].

3.2 Channel and Spatial Attention mechanisms

The primary goal of attention mechanisms in computer vision is to mimic the way human visual cognition focuses on key patterns within an image. In this research, the impact of attention mechanisms on multi-label classification tasks was assessed using two popular soft visual attention methods: Squeeze-and-Excitation (SE) [9] and Coordinate Attention (CA) [10]. Although various other attention techniques exist, such as Spatial Group-wise Enhanced Network (SGE-Net) [37], Shuffle-Attention Network

(SA-Net) [7], and Efficient Channel Attention (ECA-Net) [38], SE and CA were chosen for their roles as a baseline and state-of-the-art methods, respectively, and for their proven effectiveness in convolutional neural networks [39].

3.2.1 Squeeze-and-Excitation (Channel-based Attention)

The "squeeze" phase in the SE [9] network, generates global descriptors for each channel by pooling spatial information. This information is then utilized in the "excitation" phase to encode channel-wise dependencies and recalibrate channel-wise features. This process enhances the network's ability to represent features more effectively [9, 39]. A key feature of the Squeeze-and-Excitation mechanism is the global average pooling, which aggregates spatial information across all channels. A unique aspect of the Squeeze-and-Excitation mechanism is the use of global average pooling, which consolidates spatial information across all channels.

$$S_c = \sigma\left(W_2\,\delta\left(W_1\left(\frac{1}{HW}\sum_{i=1}^{H}\sum_{j=1}^{W}X_{ij}\right)\right)\right) \tag{1}$$

where:

- σ is the sigmoid activation function.
- δ is the ReLU activation function.
- W_1 and W_2 are the weights of two fully connected layers.
- X_{ij} denotes the input feature map value at spatial location (i, j).
- $H \times W$ are the spatial dimensions of the input feature map.

3.2.2 Coordinate Attention

The Squeeze-and-Excitation (SE) block [9] acquires global spatial information through global pooling and models inter-channel relationships but neglects the important factor of positional information. Coordinate Attention (CA) [10] overcomes this limitation by integrating positional information into channel attention, enabling the network to effectively focus on important large areas while using minimal computational resources.

The Coordinate Attention mechanism operates in two main phases: coordinate information embedding and coordinate attention generation. First, two different sizes of pooling kernels are used to process each channel, encoding information along both the horizontal and vertical axes. The outputs from these pooling layers are then concatenated and passed through a shared 1x1 convolutional transformation. Afterward, the CA mechanism splits the resulting tensor into two separate tensors. These tensors are then transformed into attention vectors aligned with the horizontal and vertical dimensions of the input X, each preserving the same number of channels.

The CA block expands its capabilities by incorporating spatial information encoding unlike traditional channel attention, which focuses mainly on recalibrating channel significance. By applying attention across both horizontal and vertical planes, CA effectively captures the positional information of target features. What sets Coordinate Attention apart from SE is its dual-axis attention mechanism – horizontal $g_c^h(i)$

and vertical $g_c^w(j)$ attention weights – that encodes spatial information along both axes (Eq. 2).

$$y_c(i,j) = x_c(i,j) \times g_c^h(i) \times g_c^w(j). \tag{2}$$

where:

- $y_c(i,j)$ is the output of the Coordinate Attention block for the c^{th} channel at position (i,j).
- $x_c(i,j)$ represents the input feature map for the c^{th} channel at position (i,j).
- $g_c^h(i)$ and $g_c^w(j)$ are the attention weights for the horizontal and vertical directions, respectively, at position (i,j).

4 Micromobility Lane Recognition Dataset (MLRD)

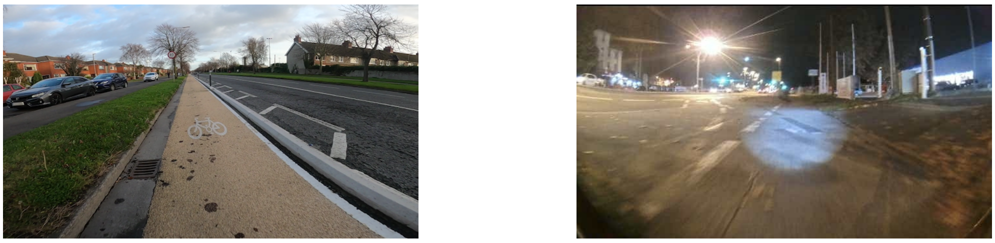
The first version of the Micromobility Lane Recognition Dataset (MLRD) is an novel multi-label classification dataset created specifically for lane recognition in micromobility safety contexts, addressing the challenges outlined in Section 1. This section outlines the key features of the dataset and describes the methodology used to approach lane recognition as a multi-label classification problem. Section 7 describes enhancements made to MLRD in light of the experimental outcomes discussed in Section 6.

The dataset consists of colour images primarily categorized into three specific classes: road, sidewalk, and bike lane. Fig. 2 and Fig. 3 show representative sample images from the dataset. These classes have been selected to meet the unique needs of micromobility vehicles, like e-scooters, ensuring efficient and safe navigation within urban settings.

(a) Road (b) Bike lane (c) Sidewalk

Fig. 2: Example images from MLRD captured with the e-scooter camera [36].

The primary motivation for developing this custom dataset is the insufficiency of existing open-source datasets for autonomous driving, such as KITTI [40], Cityscapes [41], DET [42], TuSimple [23], LLAMAS [43], CurveLanes [44], and nuScenes [45], in meeting the specific requirements of micromobility applications. While these datasets have significantly contributed to advancements in computer vision and autonomous driving research, they lack adequate representations of sidewalk and bike lane classes. The images in these datasets, typically captured from a car's



(a) Clearly visible bike lane (b) Poorly visible bike lane

Fig. 3: Contrast between a clearly delineated bike lane and a low-visibility lane.

first-person perspective, limit their relevance to micromobility scenarios that involve different perspectives and varying positions of e-bikes or e-scooters on the roads.

The images in the MLRD dataset are captured using a camera module mounted on e-scooters, capturing streets and their surroundings. These images, taken at a resolution of 640x480 (VGA), are sourced from several major cities across Europe and the United States, resulting in a dataset of 30,244 images. MLRD offers a comprehensive and diverse collection of road scenes by combining images from both online sources and the proprietary camera module, focusing specifically on roads, sidewalks, and bike lanes. Figure 4 shows the per-city distribution of images, illustrating geographic diversity and the relative sampling across regions.

This targeted approach is designed to facilitate the development of more accurate and efficient models for lane detection and classification, ultimately contributing to safer and more intelligent urban mobility solutions. Additionally, the privacy of individuals and vehicle licence plates was ensured by blurring out faces and licence plates in the images. The MLRD dataset has been made publicly available.

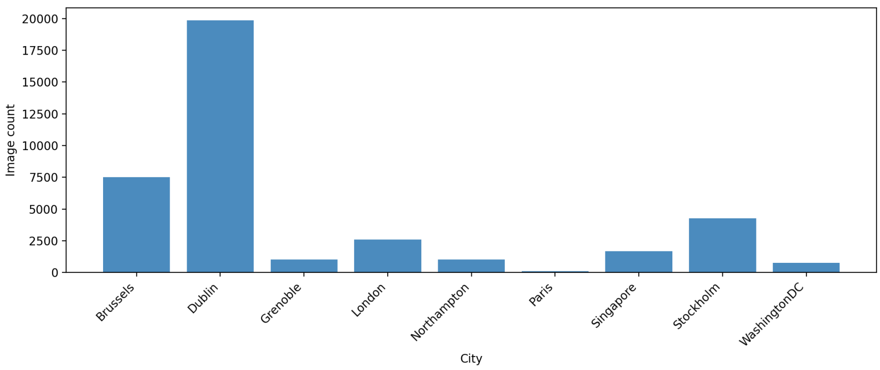


Fig. 4: Per-city distribution of MLRD dataset images illustrating broad geographic diversity across European and U.S. urban environments.

Due to the diverse structures of roads and sidewalks, as well as the varying positions of these segments (whether adjacent or separated), precise annotation of frames was a

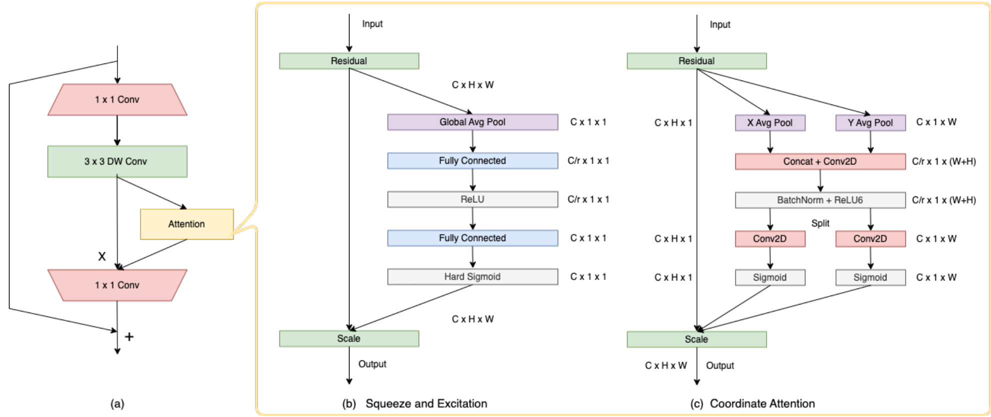


Fig. 5: An illustration of the integration of channel and channel plus spatial attention blocks in the inverted residual block (a) present in MobileNetV2 [12] and MobileNetV3 [11] architectures, as utilized in our experiments [36].

challenging task. Particularly, the bike lanes are very different in different cities. For example, Figure 3a shows some examples from Dublin City where there are clear bike lane markings with borders and distinct colouring; however, there are many scenarios, as illustrated in Figure 3b, where the bike lanes are subtly distinct and contain very little visual information to distinguish that class from the road or sidewalk. Therefore, priority in labelling was given to the road and bike lane classes. Images that clearly display a road or bike lane were labelled as 1 for that class and 0 for the other. In cases where the bike lane is part of the road segment, the image is annotated as both road and bike lane. Conversely, if an image clearly depicts a sidewalk or lacks clear road or bike lane areas, it is annotated with 0 for both the road and bike lane labels. The MLRD dataset includes a total of 16,759 samples labelled as "road", 5,218 samples as"bike lane", and 12,510 samples for the indirect class "sidewalk", where both the "road" and "bike lane" labels are set to zero. Although the dataset is slightly imbalanced, this issue was addressed using an appropriate loss function, as discussed in Section 5.

5 Experimental Methodology

This section details the methodology, model deployment, and implementation specifics of the comparative analysis conducted. The experiments evaluated the impact of channel and spatial attention on the compact MobileNets (with the width multiplier $\alpha = 0.1$) for classification by comparing the performance of five different variants: Standard MobileNetV2 [12], MobileNetV2 with channel attention, MobileNetV2 with channel and spatial attention, Standard MobileNetV3 [11] with channel attention, and MobileNetV3 with channel and spatial attention on the MLRD dataset.

5.1 Experimental setup and hyper-parameter details

Experiments were performed using the TensorFlow framework, with the Weights and Biases MLOps tool employed to track all metrics during training. The training was conducted on an NVIDIA GEFORCE RTX 4090 GPU, with an input image resolution of 224 x 224 and a batch size of 32. The Adam optimizer was initialized with a learning rate (LR) of 0.001. The "ReduceLROnPlateau" learning rate scheduler was configured to monitor validation loss, and "ModelCheckpoint" was implemented to save the best model based on the lowest validation loss. The minimum learning rate was set to 1e-6, with a reduction factor of 0.1 and a patience parameter of 10 epochs. All models were trained from scratch on the MLRD without using any pre-trained weights. Minimal image augmentations, such as horizontal flips and brightness adjustments within the range of 0.2-0.5, were applied during training, while augmentations like vertical flips or rotations were avoided to maintain the integrity of the first-person micromobility rider perspective in the images. To address the class imbalance in MLRD and achieve the multi-label classification objective, the Binary Focal Cross entropy (BFCE) loss function [46] was utilized, with a weight balancing factor (α) of 0.25 and a focusing parameter (γ) of 2.0. All models were trained for 80 epochs.

5.2 Network architecture

In this study, it was empirically determined that approximately 100K trainable parameters, combined with an input image resolution of 224x224, enabled the successful deployment of the model on the target platform. A detailed description of the technical specifications for the deployment platform is provided in Section 5.3.

The MobileNetV2 [12] and MobileNetV3 [11] based model architectures, including the baseline used in this research, were derived from the official *Keras* implementation. In accordance with the official guidelines for Squeeze-and-Excitation (SE)[9] and Coordinate Attention (CA)[10], the SE blocks were strategically positioned immediately after the depthwise convolution layers within the bottleneck modules of the MobileNet architectures, as illustrated in Figure 5. This placement allowed the SE blocks to recalibrate the features extracted by depthwise separable convolutions before their projection through pointwise convolutions (1x1 convolutions) into a higher-dimensional space.

The width multiplier parameter, which controls the number of channels in the bottleneck layers, was fixed at 0.1 for all the MobileNet model variants in this research. This adjustment did not change the network's depth. The channel reduction ratio for the squeeze operation in the SE block was set to 16, while for the channel attention phase in CA, it was established at 32.

To meet the memory constraints of the target hardware platform, particularly for MobileNetV3, the channels in the bottleneck blocks were modified while maintaining the original ratio and preserving the integrity of the original model architecture. Specifically, the number of output channels was halved, and the expansion factor in each bottleneck block was reduced to one-fourth of their original values. The expansion factor, which is used to increase the number of channels in the input feature

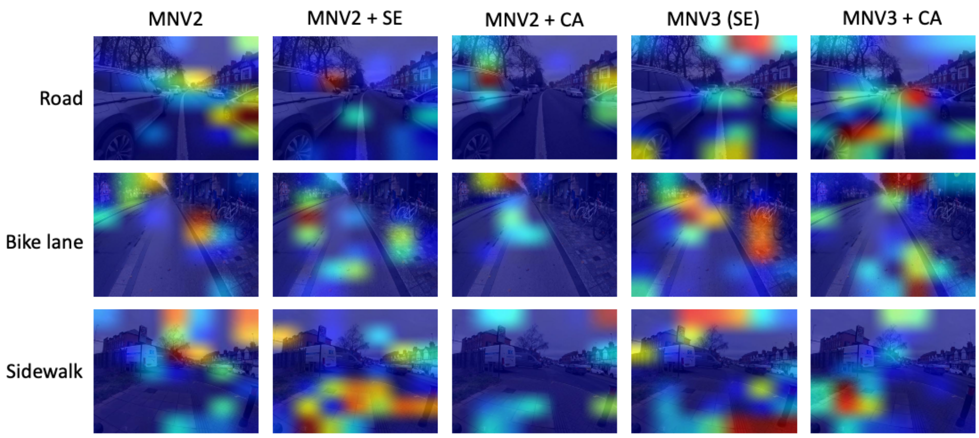


Fig. 6: A comparative Grad-CAM visualizations illustrating distinct behavioral patterns of different MobileNet model variants on the MLRD dataset [36].

map before applying a depthwise separable convolution, allows the network to capture more complex features in a higher-dimensional space while maintaining overall computational efficiency.

Additionally, the neuron count in the final Fully Connected layer was reduced from 1280 to 320 to align the model with the computational limitations of the deployment platform.

5.3 Deployment on the target platform

The platform chosen for our project is the STM32H743VI, a compact microcontroller unit measuring 1.40 in x 1.75 in. This unit features a 32-bit Arm Cortex-M7 processor running at 480 MHz, with 1 MB of static RAM and 2 MB of flash memory. In addition to its small size, the unit consumes less than 150 mA, making it well-suited for the limited power requirements of micromobility vehicles. As illustrated in Figure 1, the final step in our process involves converting the models to *TFLite* format for deployment on the target platform. However, the float32 compact models mentioned earlier exceeded the required size limits, preventing successful deployment. To address this, the model size was reduced by converting the weights and activations from *float32* to *int8* precision, using Post-Training Quantization [47].

Due to these constraints, only the compact MobileNetV2 model (baseline) could be deployed, as the platform's firmware did not support certain tensor operations found in the MobileNetV3 architecture at the time the experiments were conducted. The deployed compact MobileNetV2 model had the model size of 250 KB and exhibited an inference latency corresponding to approximately 10–12 FPS. In future, as more operations are supported for this MCU, it will be possible to compile and run these models directly on the platform. Therefore, our conclusions are based on experiments carried out on a local desktop machine, with the expectation that there may be slight differences in behaviour when deployed on the actual platform, primarily due to the quantization process used to compress the model.

6 Results and Discussions

Table 1 presents the performance metrics for the standard large MobileNetV2 (MNV2) and MobileNetV3 (MNV3) base models on the MLRD dataset. These models, having a much larger number of parameters compared to the compact versions shown in Table 2, exhibit a notable difference in performance. The main goal of incorporating attention mechanisms into the compact models is to reduce this performance gap while maintaining model efficiency and avoiding a significant increase in computational cost.

Table 1: Classification results between standard (Large) MobileNets on MLRD [36].

	Precision		Recall		F1 Score		Weighted Avg F1	Trainable param.
	Road	Bike lane	Road	Bike lane	Road	Bike lane		
MobileNetV2	0.96	0.97	0.90	0.86	0.93	0.92	0.93	2.26M
MobileNetV3 Large	0.95	0.94	0.92	0.87	0.94	0.91	0.93	2.99M

Table 2: Classification results between Attention-Based and Non-Attention-Based compact versions (compact) of MobileNets on MLRD [36]. Bold values indicate significance, showing the highest overall scores for MobileNetV2 + CA and parameter-efficient competitive performance for MobileNetV3 + CA.

Models (*compact*)	Precision		Recall		F1 Score		Weighted Avg F1	Trainable param.
	Road	Bike lane	Road	Bike lane	Road	Bike lane		
MobileNetV2 (Baseline)	0.86	0.86	0.89	0.80	0.88	0.83	0.86	95.87K
MobileNetV2 + SE	0.87 ↑1%	0.88 ↑2%	0.89	0.79 ↓1%	0.88	0.83	0.87 ↑1%	106.79K
MobileNetV2 + CA	**0.94 ↑8%**	**0.93 ↑7%**	0.83 ↓6%	0.79 ↓1%	0.88	**0.85 ↑2%**	**0.87 ↑1%**	124.21K
MobileNetV3 (SE)	0.83 ↓3%	0.77 ↓9%	0.74 ↓15%	0.84 ↑4%	0.78 ↓10%	0.80 ↓3%	0.79 ↓7%	109.82K
MobileNetV3 + CA	0.83 ↓3%	**0.86**	0.85 ↓4%	**0.81 ↑1%**	0.84 ↓4%	0.83	**0.84** ↓2%	88.74K

In the comparative analysis of MobileNetV2 and MobileNetV3 architectures, integrated with Squeeze-and-Excitation (SE) and Coordinate Attention (CA) mechanisms respectively, the experimental results underscore the importance of channel and spatial information for lane recognition applications. These findings are particularly significant given the challenges of working with relatively low-resolution images in a resource-constrained environments, where capturing maximum features with minimal computational load is essential. The MNV2 model, integrated with Coordinate Attention, showed notable improvements in precision for both road and bike lane classification, achieving an approximately 8% increase in precision and a 1% improvement in the average F1 score compared to the baseline model. This boost in precision is vital for reducing false positives, which is critical for the reliability and robustness of micromobility safety systems.

However, this configuration also resulted in the highest trainable parameter count, adding approximately 29K parameters, which translates to about 100KB of additional memory usage, reflecting a 29.55% increase from the baseline. While this setup may offer optimal performance, its applicability in low-resource environments is limited due to the increased computational overhead. A decline in the recall metric, particularly for the "road" class, was observed for this model, indicating potential overfitting. This suggests that the model may have become overly specialized in identifying certain features, likely those associated with the "bike lane" class, at the expense of its generalization capabilities. This underscores the importance of finding the right balance between model complexity and its ability to perform consistently across a variety of real-world conditions.

On the contrary, the MNV2 model, integrated with Squeeze-and-Excitation attention (channel attention), added approximately 11K parameters but led to only a very slight improvement in precision for both classes, along with a marginal increase in the overall F1 score. Nevertheless, this improvement still highlights the effectiveness of channel-based attention mechanisms, such as Squeeze-and-Excitation, in enhancing model performance, as the recall metric remained almost unchanged for both classes compared to the baseline. However, its ability to fully capture the spatial complexities necessary for tasks like lane recognition remains limited compared to MobileNetV2 with Coordinate Attention.

For MNV3-based models, the variant with Coordinate Attention demonstrated consistently strong performance across all metrics while maintaining a smaller parameter footprint compared to the MNV2 model, making it more suitable for low-spec microcontrollers. Notably, the standard MNV3 model with SE attention was outperformed by the model integrated with Coordinate Attention, likely due to SE attention's limitations in capturing the spatial complexities essential for accurate lane detection.

The integration of Coordinate Attention to the MNV2 architecture improved performance over both the baseline and channel attention-based variants. This improvement is largely attributed to its dual emphasis on spatial and channel-wise feature interdependencies. Such an approach is critical for distinguishing visually similar classes, like roads and bike lanes, where spatial positioning plays a key role. The limited effectiveness of SE attention in these image classification experiments can be considered as a result of its focus on channel-wise feature recalibration, which lacks the spatial sensitivity required for tasks that demand a deeper understanding of positional context.

The impact of these models extends beyond micromobility safety, holding considerable potential for autonomous driving solutions, where accurate and efficient lane recognition is essential. Deploying these compact yet powerful models on low-spec microcontrollers opens up a promising avenue for creating cost-efficient, high-performance autonomous navigation and driving safety systems, providing a resource-efficient alternative to more computationally demanding segmentation models. This strategy enables the development of accurate and economically feasible safety systems, particularly suited to environments requiring fast inference and low computational overhead.

Figure 6 shows a Grad-CAM [48] analysis of all the compact model variants executed on a few sample images from the MLRD. This visualization highlights the unique behaviour of each model. However, no consistent pattern emerges regarding the localisation of the complex differentiating features of the classes of interest to determine the final predictions.

7 Micromobility Lane Recognition Dataset V2 (MLRDv2)

A limitation of the first version of MLRD was the lack of variety of environment conditions and riding scenarios. MLRDv2 comprises 29,486 colour images captured from a first-person perspective by mounting a mini camera on the front stem of an e-scooter. This extension of the original MLRD features a broader range of curated frames that add diversity and complexity. Table 3 shows the distribution of samples across environmental conditions (daytime, nighttime, rainy, and rainy nighttime), highlighting the intentional inclusion of challenging and realistic scenarios.

Table 3: Environmental condition distribution in MLRDv2. Each row lists the number of dataset samples and their share of the dataset.

Condition	Dataset samples	Overall %
Day time	24,979	63.49
Night time	8,050	20.46
Rainy	5,043	12.82
Rainy & night	1,271	3.23

The images in MLRDv2 are categorized into key classes such as “road” and “bike lane”, where micromobility vehicles are permitted to operate, and an indirect “sidewalk” class, where riding is prohibited. The labelling guidelines adhered to in MLRDv2 are consistent with those used in the original MLRD, as detailed in Section 4.

Fig. 7: MLRDv2 example images captured in rainy weather.

A significant enhancement in MLRDv2 is the inclusion of more naturalistic scenes, particularly those captured under challenging conditions such as rainy weather, nighttime, and the especially complex scenario of rainy nights. Figure 2 contrasts this with

Table 4: Classification results between standard (Large) and compact MobileNets on MLRDv2.

Models	Precision		Recall		F1 Score		Weighted Avg F1	Trainable param.
	Road	Bike lane	Road	Bike lane	Road	Bike lane		
MNV2	0.98	0.97	0.99	0.94	0.98	0.96	0.98	2.226M
MNV3Small	0.98	0.96	0.98	0.93	0.98	0.94	0.97	1.51M
MNV3Large	0.98	0.96	0.98	0.93	0.98	0.94	0.97	4.2M
CompMNV2	0.87	0.70	0.87	0.53	0.87	0.60	0.82	0.88M
CompMNV3	0.81	0.52	0.77	0.46	0.79	0.49	0.74	1.11M

samples from the original MLRD, which predominantly features images taken in clear weather during daylight hours. For a model to be practically useful, it must perform well under less ideal conditions, such as when rain causes water droplets to obscure the camera lens. Figure 7 illustrates some of these challenging rainy scenes from MLRDv2.

Fig. 8: MLRDv2 example images captured at night time.

Nighttime scenes are notoriously difficult due to low visibility and minimal light exposure. Figure 8 showcases examples from MLRDv2 that depict these nighttime challenges. Furthermore, recognizing that both rain and nighttime conditions can simultaneously impact visibility, MLRDv2 also includes images captured during rainy nights, offering a more comprehensive training set for real-world applications. The example images captured at rainy nights can be seen in Figure 9

Fig. 9: MLRDv2 example images captured in rainy weather at night time.

MLRDv2 offers a more varied and realistic set of scenes compared to its predecessor, making it a valuable resource for developing robust safety solutions for micromobility.

7.1 Dataset Validation

This section describes the implementation details for assessing the MLRDv2 dataset. Five variants of the MobileNet model were used , including three standard MobileNet versions and two compact models. Specifically, the models used were MobileNet V2 (MNV2), MobileNet V3 Small (MNV3Small), and MobileNet V3 Large (MNV3Large), all with a width multiplier of 1.0, as well as compact versions of MNV2 and MNV3Small with width multipliers of 0.1 and 0.25, respectively to maintain the total trainable parameters as mentioned in Section 5.2. All experiments were conducted using the PyTorch framework. The standard MobileNet models were initialized with pretrained ImageNet weights, while the compact models were trained from scratch, as the pretrained ImageNet weights are not available for width multipliers of 0.1 and 0.25 for MobileNetV2 and MobileNetV3. All models were trained for 100 epochs with a batch size of 64 and a learning rate of 1e-4. For data augmentation, all images were resized to 224x224 pixels, normalized using the ImageNet mean and standard deviation values, and 50% of the images were flipped horizontally. The dataset was split into 70% for training, with 15% each allocated for validation and testing. We utilized the Adam optimizer with default momentum and weight decay settings, and binary cross-entropy loss as the loss function.

7.2 Results and Discussions

Table 4 presents the performance metrics on the MLRDv2 test set for both the standard and compact models trained on MLRDv2. The first three rows display the precision, recall, and F1 scores for the standard MobileNet models. These models demonstrate consistently high performance metrics on the test set, indicating that they successfully learned the relevant features from the MLRDv2 dataset.

8 Conclusions and Future Work

Cities worldwide are advancing toward eco-friendly urban transportation systems, such as e-scooters and e-bikes, to promote sustainable mobility and reduce traffic, noise, and pollution. Understanding the behavior of e-mobility users is essential for effective regulation and governance. However, current GPS and LIDAR systems have limitations, highlighting the need for computer vision-based solutions. A significant challenge is the lack of dedicated image datasets for micromobility safety, as existing datasets for autonomous driving differ substantially in perspective and operating environments. Most available datasets focus on objects like pedestrians, vehicles, and obstacles, leaving a notable gap in lane recognition data. To address this, we developed a novel lane recognition multi-label image classification dataset specifically designed for micromobility applications.

Our experiments with MLRD have shown that integrating attention mechanisms into compact CNN models, such as MobileNetV2 and MobileNetV3, can improve precision and overall performance, but not without challenges. For instance, the MobileNetV3 model with both channel and spatial attention demonstrated impressive performance, closely matching that of the baseline model while using significantly fewer parameters. Conversely, the MobileNetV2 model with channel and spatial attention improved precision but experienced a decline in recall for the "road" class. Similarly, the MobileNetV3 model with channel attention showed an overall performance drop compared to the baseline, suggesting potential overfitting. These results indicate that while attention mechanisms can boost model accuracy, their integration into compact models must be carefully managed to balance the benefits against the risks of overfitting and increased complexity.

To further improve the dataset, we recognized the growing demand in micromobility and the need for a robust dataset. As a result, we added more realistic frames from various outdoor infrastructures, adverse weather conditions, and nighttime environments with low visibility, making the dataset more comprehensive and representative of real-world scenarios. This enhanced version, MLRDv2, proved more challenging and thorough. Standard MobileNet models continued to perform well, while compact models—especially MobileNetV3 with channel attention—displayed trends of overfitting.

In conclusion, our dataset, MLRDv2, serves as a valuable resource for assessing the effectiveness of MobileNet models with channel and spatial attention mechanisms in improving lane recognition accuracy. These attention mechanisms show promise for compact models in micromobility, particularly with MobileNetV2 variants that achieved higher F1 scores with minimal increases in parameter count. However, the challenges of model compression and deployment in micromobility environments can sometimes diminish the potential improvements. Moving forward, we aim to conduct a more comprehensive comparison across various model architectures and deploy these models on low-spec target platforms to assess real-world performance. Additionally, we plan to explore more efficient model optimization techniques, such as structured weight pruning and quantization-aware training, to enhance both performance and hardware compatibility.

ACKNOWLEDGEMENTS

We would like to express our gratitude to Luna Systems for their invaluable support throughout the course of this research. Their generosity in providing us with the essential dataset and the microcontroller platform was instrumental in the successful completion of our experiments.

AUTHOR CONTRIBUTIONS

CK contributed to the design of the study, data curation, design of experiments, actual implementation, running evaluations, analysing and visualizing results, writing—original draft, and proofreading—original draft. PC contributed to data curation, design of experiments, analysing and visualizing results, and proofreading—original draft. BC was responsible for dataset creation and proofreading—original draft. AF contributed to dataset creation and proofreading—original draft. ML was involved in analysing and visualizing results, writing—original draft, and proofreading—original draft. SL contributed to the design of the study, design of experiments, analysing and visualizing results, writing—original draft, and proofreading—original draft.

FUNDING

This research was conducted with the financial support of Research Ireland (12/RC/2289_P2), at the Research Ireland Insight Centre for Data Analytics at Dublin City University, and Luna Systems.

AVAILABILITY OF DATA AND MATERIALS

Our code scripts and dataset are publicly accessible at: https://github.com/Luna-Scooters/Compact-Attention-based-CNNs-on-MLRD.

DECLARATIONS

Competing interests

On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

[1] Cicchino, J.B., Kulie, P.E., McCarthy, M.L.: Severity of e-scooter rider injuries associated with trip characteristics. Journal of safety research **76**, 256–261 (2021)

[2] Fox, A., Kumar, B.V., Chen, J., Bai, F.: Multi-lane pothole detection from crowdsourced undersampled vehicle sensor data. IEEE Transactions on Mobile Computing **16**(12), 3417–3430 (2017)

[3] Xing, Y., Lv, C., Chen, L., Wang, H., Wang, H., Cao, D., Velenis, E., Wang, F.-Y.: Advances in vision-based lane detection: Algorithms, integration, assessment, and perspectives on acp-based parallel vision. IEEE/CAA Journal of Automatica Sinica **5**(3), 645–661 (2018)

[4] Zakaria, N.J., Shapiai, M.I., Ghani, R.A., Yasin, M., Ibrahim, M.Z., Wahid, N.: Lane detection in autonomous vehicles: A systematic review. IEEE Access (2023)

[5] Zhu, X., Cheng, D., Zhang, Z., Lin, S., Dai, J.: An empirical study of spatial attention mechanisms in deep networks. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 6688–6697 (2019)

[6] Fu, J., Liu, J., Jiang, J., Li, Y., Bao, Y., Lu, H.: Scene segmentation with dual relation-aware attention network. IEEE Transactions on Neural Networks and Learning Systems **32**(6), 2547–2560 (2020)

[7] Zhang, Q.-L., Yang, Y.-B.: Sa-net: Shuffle attention for deep convolutional neural networks. In: ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 2235–2239 (2021). IEEE

[8] GM, V., Pereira, B., Little, S.: Urban footpath image dataset to assess pedestrian mobility. In: Proceedings of the 1st International Workshop on Multimedia Computing for Urban Data, pp. 23–30 (2021)

[9] Hu, J., Shen, L., Sun, G.: Squeeze-and-excitation networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7132–7141 (2018)

[10] Hou, Q., Zhou, D., Feng, J.: Coordinate attention for efficient mobile network design. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 13713–13722 (2021)

[11] Howard, A., Sandler, M., Chu, G., Chen, L.-C., Chen, B., Tan, M., Wang, W., Zhu, Y., Pang, R., Vasudevan, V., *et al.*: Searching for mobilenetv3. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 1314–1324 (2019)

[12] Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., Chen, L.-C.: Mobilenetv2: Inverted residuals and linear bottlenecks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4510–4520 (2018)

[13] Sabri, K., Djilali, C., Bilodeau, G.-A., Saunier, N., Bouachir, W.: Detection of micromobility vehicles in urban traffic videos. Proceedings of the 21st Conference on Robots and Vision (2024)

[14] Apurv, K., Tian, R., Sherony, R.: Detection of e-scooter riders in naturalistic scenes. arXiv preprint arXiv:2111.14060 (2021)

[15] Chen, D.: Object Detection for E-scooters

[16] Su, K., Tomioka, Y., Zhao, Q., Liu, Y.: Yolic: An efficient method for object localization and classification on edge devices. Image and Vision Computing **147**, 105095 (2024)

[17] Nienaber, S., Booysen, M., Kroon, R.: Detecting potholes using simple image processing techniques and real-world footage. (2015). Southern African Transport Conference

[18] Nienaber, S., Kroon, R., Booysen, M.J.: A comparison of low-cost monocular vision techniques for pothole distance estimation. In: 2015 IEEE Symposium Series on Computational Intelligence, pp. 419–426 (2015). IEEE

[19] Erdelić, M., Erdelić, T., Carić, T.: Dataset for multimodal transport analytics of smartphone users - collecty. Data in Brief **50**, 109481 (2023)

[20] Kim, E., Ryu, H., Oh, H., Kang, N.: Safety monitoring system of personal mobility driving using deep learning. Journal of Computational Design and Engineering

9(4), 1397–1409 (2022)

[21] Zhang, L., Jiang, F., Kong, B., Yang, J., Wang, C.: Real-time lane detection by using biologically inspired attention mechanism to learn contextual information. Cognitive Computation **13**, 1333–1344 (2021)

[22] Woo, S., Park, J., Lee, J.-Y., Kweon, I.S.: Cbam: Convolutional block attention module. In: Proceedings of the European Conference on Computer Vision (ECCV), pp. 3–19 (2018)

[23] Chang, D., Chirakkal, V., Goswami, S., Hasan, M., Jung, T., Kang, J., Kee, S.-C., Lee, D., Singh, A.P.: Multi-lane detection using instance segmentation and attentive voting. In: 2019 19th International Conference on Control, Automation and Systems (ICCAS), pp. 1538–1542 (2019). IEEE

[24] Aly, M.: Real time detection of lane markers in urban streets. In: 2008 IEEE Intelligent Vehicles Symposium, pp. 7–12 (2008). IEEE

[25] Li, J., Jiang, F., Yang, J., Kong, B., Gogate, M., Dashtipour, K., Hussain, A.: Lane-deeplab: Lane semantic segmentation in automatic driving scenarios for high-definition maps. Neurocomputing **465**, 15–25 (2021)

[26] Pan, X., Shi, J., Luo, P., Wang, X., Tang, X.: Spatial as deep: Spatial cnn for traffic scene understanding. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 32 (2018)

[27] Lee, M., Lee, J., Lee, D., Kim, W., Hwang, S., Lee, S.: Robust lane detection via expanded self attention. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pp. 533–542 (2022)

[28] Yu, F., Chen, H., Wang, X., Xian, W., Chen, Y., Liu, F., Madhavan, V., Darrell, T.: Bdd100k: A diverse driving dataset for heterogeneous multitask learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 2636–2645 (2020)

[29] Yao, Z., Chen, X., et al.: Efficient lane detection technique based on lightweight attention deep neural network. Journal of Advanced Transportation **2022** (2022)

[30] Hou, Y., Ma, Z., Liu, C., Loy, C.C.: Learning lightweight lane detection cnns by self attention distillation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 1013–1021 (2019)

[31] Hosseini, S.R., Taheri, H., Teshnehlab, M.: Enet-21: An optimized light cnn structure for lane detection. arXiv preprint arXiv:2403.19782 (2024)

[32] Khan, M.A.-M., Haque, M.F., Hasan, K.R., Alajmani, S.H., Baz, M., Masud, M., Nahid, A.-A.: Lldnet: a lightweight lane detection approach for autonomous cars

using deep learning. Sensors **22**(15), 5595 (2022)

[33] Hanhirova, J., Kämäräinen, T., Seppälä, S., Siekkinen, M., Hirvisalo, V., Ylä-Jääski, A.: Latency and throughput characterization of convolutional neural networks for mobile computer vision. In: Proceedings of the 9th ACM Multimedia Systems Conference, pp. 204–215 (2018)

[34] Elsken, T., Metzen, J.H., Hutter, F.: Neural architecture search: A survey. The Journal of Machine Learning Research **20**(1), 1997–2017 (2019)

[35] Sanchez-Iborra, R., Skarmeta, A.F.: Tinyml-enabled frugal smart objects: Challenges and opportunities. IEEE Circuits and Systems Magazine **20**(3), 4–18 (2020)

[36] Kaundanya, C., Cesar, P., Cronin, B., Fleury, A., Liu, M., Little, S.: Using Attention Mechanisms in Compact CNN Models for Improved Micromobility Safety Through Lane Recognition. In: Proceedings of the 10th International Conference on Vehicle Technology and Intelligent Transport Systems - Volume 1: VEHITS, pp. 88–98 (2024). INSTICC

[37] Li, X., Hu, X., Yang, J.: Spatial group-wise enhance: Improving semantic feature learning in convolutional networks. arXiv preprint arXiv:1905.09646 (2019)

[38] Wang, Q., Wu, B., Zhu, P., Li, P., Zuo, W., Hu, Q.: Eca-net: Efficient channel attention for deep convolutional neural networks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 11534–11542 (2020)

[39] Guo, M.-H., Xu, T.-X., Liu, J.-J., Liu, Z.-N., Jiang, P.-T., Mu, T.-J., Zhang, S.-H., Martin, R.R., Cheng, M.-M., Hu, S.-M.: Attention mechanisms in computer vision: A survey. Computational visual media **8**(3), 331–368 (2022)

[40] Geiger, A., Lenz, P., Stiller, C., Urtasun, R.: Vision meets robotics: The kitti dataset. The International Journal of Robotics Research **32**(11), 1231–1237 (2013)

[41] Cordts, M., Omran, M., Ramos, S., Rehfeld, T., Enzweiler, M., Benenson, R., Franke, U., Roth, S., Schiele, B.: The cityscapes dataset for semantic urban scene understanding. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3213–3223 (2016)

[42] Cheng, W., Luo, H., Yang, W., Yu, L., Chen, S., Li, W.: Det: A high-resolution dvs dataset for lane extraction. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, pp. 0–0 (2019)

[43] Behrendt, K., Soussan, R.: Unsupervised labeled lane markers using maps. In: Proceedings of the IEEE/CVF International Conference on Computer Vision

Workshops, pp. 0–0 (2019)

[44] Xu, H., Wang, S., Cai, X., Zhang, W., Liang, X., Li, Z.: Curvelane-nas: Unifying lane-sensitive architecture search and adaptive point blending. In: Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XV 16, pp. 689–704 (2020). Springer

[45] Caesar, H., Bankiti, V., Lang, A.H., Vora, S., Liong, V.E., Xu, Q., Krishnan, A., Pan, Y., Baldan, G., Beijbom, O.: nuscenes: A multimodal dataset for autonomous driving. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 11621–11631 (2020)

[46] Lin, T.-Y., Goyal, P., Girshick, R., He, K., Dollár, P.: Focal loss for dense object detection. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2980–2988 (2017)

[47] Zhang, J., Zhou, Y., Saab, R.: Post-training quantization for neural networks with provable guarantees. SIAM Journal on Mathematics of Data Science **5**(2), 373–399 (2023)

[48] Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D.: Grad-cam: Visual explanations from deep networks via gradient-based localization. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 618–626 (2017)