Weather and Meteorological Optical Range Classification for Autonomous Driving

Celso Pereira, Ricardo P. M. Cruz, João N. D. Fernandes, João Ribeiro Pinto, and Jaime S. Cardoso, *Senior Member, IEEE*

***Abstract*—Weather and meteorological optical range (MOR) perception is crucial for smooth and safe autonomous driving (AD). This article introduces two deep learning-based architectures, employing early and intermediate sensor fusion and multi-task strategies, designed for concurrent weather and MOR classification in AD. Extensive experiments employing the publicly available FogChamber dataset demonstrate that the proposed early fusion architecture, characterized by its lightweight design and simplicity, achieves an accuracy of 98.88% in weather classification and 89.77% in MOR classification, with a competitive memory allocation of 5.33 megabytes (MB) and an inference time of 2.50 milliseconds (ms). In contrast, the proposed intermediate fusion architecture prioritizes performance, achieving higher accuracies of 99.38% in weather classification and 91.88% in MOR classification. However, it requires a more substantial memory allocation of 54.06 MB and exhibits a longer inference time of 15.55 ms. Compared to other state-of-the-art architectures, the proposed methods present a competitive balance between accuracy performance, inference time, and memory allocation, which are crucial parameters for enabling autonomous driving.**

***Index Terms*—Weather classification, Meteorological optical range (MOR) classification, Deep learning, Multi-task learning, Single-task learning, Multi-modal, RGB Camera, LiDAR, Image entropy.**

I. Introduction

Autonomous vehicles (AVs) rely on various sensors, such as cameras and LiDAR, to perceive the surrounding environment and make informed decisions. Nonetheless, adverse weather conditions can potentially disrupt the effectiveness of these sensors through the scattering and absorption of emitted and reflected signals, or by degenerating image contrast. This can ultimately result in degraded or inaccurate perception data, making it challenging for AVs to accurately detect and recognize obstacles, navigate lanes, and make appropriate driving decisions. Therefore, weather awareness is paramount for AVs as it allows the vehicle's behavior to be adjusted to the weather conditions. According to data compiled by the National Highway Traffic Safety Administration[1] (NHTSA), over 5 million road accidents occurred in 2020. Around 14% of these accidents were weather-related. In this article, we focus on local weather classification, as traditional weather forecasts do not offer the density and resolution required to accurately reflect the meteorological effects experienced by an individual AV and typically rely on Internet connectivity.

In the context of autonomous driving (AD), meteorological optical range (MOR) holds significant relevance for assessing safe driving conditions. MOR is defined as the distance through the atmosphere needed to reduce the luminous flux in a collimated beam from an incandescent lamp, at a color temperature of 2700 K, to 5% of its initial value.[2] This measure of visibility is important in many applications, including aviation and ground transportation [1], [2]. MOR is affected by adverse weather conditions such as fog and rain, as well as other factors such as pollution and dust. As a reference, fog represents an atmospheric opacity with a visibility of less than 1000 meters and, according to Chaabani et al.[3], can be considered dense when the MOR falls below 40 meters and thick when the MOR is between 40 meters and 200 meters. Optical visiometers, such as transmissometers and scatterometers, are commonly used for accurate MOR measurement at long distances; however, these devices are bulky and therefore not suitable for AD, so other sensor technologies, such as camera and LiDAR, should be employed.

Combining data from multiple sensors has been shown to improve AD performance [4], [5], [6]. This combination allows the individual limitations associated with each sensor to be mitigated. Therefore, multi-modal fusion strategies, such as early fusion and intermediate fusion, are crucial to take advantage of the different modalities and optimize overall performance.

Given that weather and MOR are closely related, we propose two multi-modal multi-task deep learning-based architectures that leverage data from both camera and LiDAR for weather

Manuscript received 26 March 2024; revised 4 April 2024; accepted 6 April 2024. Date of publication 10 April 2024; date of current version 18 June 2025. This work was supported in part by Portuguese National Funds, through AICEP, E. P. E. and in part by the Innovation and Digital Transition Program under Grant COMPETE2030, and in part by Project ATLAS – Trusted Autonomous Navigation, with Funding Reference 01/RPA/2022-C679908640-00009887. *(Corresponding author: Celso Pereira.)*

Celso Pereira, Ricardo P. M. Cruz, João N. D. Fernandes, and Jaime S. Cardoso are with the INESC TEC, 4200-465 Porto, Portugal, and also with the Faculty of Engineering of the University of Porto, 4200-465 Porto, Portugal (e-mail: up202200546@up.pt; rpcruz@fe.up.pt; joao.n.fernandes@inesctec.pt; jsc@fe.up.pt).

João Ribeiro Pinto was with Bosch CarMultimedia Portugal, S.A., 4701-970 Braga, Portugal. He is now with Vision-Box - Soluções de Visão por Computador, S.A., 2790-204 Carnaxide, Portugal (e-mail: jtrpinto@gmail.com).

The code presented in this research is available at https://github.com/CelsoPereira1/AutoWeatherPerception.

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TIV.2024.3387113.

Digital Object Identifier 10.1109/TIV.2024.3387113

[1] NHTSA | Traffic Safety Facts 2020. Available: https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/813369

[2] IALA | Meteorological Optical Range. Available: https://www.iala-aism.org/wiki/dictionary/index.php/Meteorological_Optical_Range

and MOR classification for AD. The first architecture adopts early fusion in order to retain rich information and minimize computational complexity. In contrast, the second architecture uses intermediate fusion and capitalizes on the self-attention mechanism of transformers to fuse the global context from both camera and LiDAR modalities. The use of both camera and LiDAR sensors is based on their frequent recognition as the most susceptible sensors to adverse weather conditions in AD [7]. The proposed objectives are addressed in the form of two classification tasks, for weather (in our experiments, fog and rain), and MOR (which involves multi-class classification with three distinct classes in our experiments).

Multi-task aims to improve data efficiency over the single-task, so a direct comparison of the two tasks is made for each architecture regarding accuracy performance, inference time, and memory allocation.

A. Contributions

- Introduction of two multi-modal multi-task deep learning-based architectures:
  - Focus on maximizing efficiency while achieving state-of-the-art performance in weather and MOR classification within the context of AD;
  - Significant advances in concurrently addressing these two crucial tasks.
- Conduction of three detailed ablation studies:
  - Evaluation of various architectures for concurrent weather and MOR classification using the FogChamber dataset;
  - Comparison between these architectures and their corresponding single-task equivalents;
  - Analysis of the effects of two optimization techniques: multi-adaptive and fixed loss weighting;
  - Insights that contribute to understanding the trade-offs of the explored architectures.

B. Article Structure

Following the Introduction, the article is organized as follows: Section II describes the state-of-the-art and previous contributions. The methodology is detailed in Section III. In Section IV, the implementation process is described and the experimental results are reported. Finally, Section V presents the conclusions, and Section VI outlines future research directions.

II. Related Work

Past contributions are highlighted for each task in isolation since, as far as we are aware, no research article in the literature simultaneously estimates weather and MOR. Furthermore, an overview of previous research works incorporating attention mechanisms in AD is provided.

A. Weather Estimation

In the literature, numerous research articles validate the decline in camera and LiDAR perception during adverse weather conditions [8], [9], [10], [11], [12], with several articles also employing these sensors for weather estimation. In the research by Sebastian et al. [13], a deep convolutional neural network named RangeWeatherNet was introduced for weather and road condition estimation. The LiDAR range view projection (range, x, y, z, pulse intensity) serves as input to a convolutional encoder based on the DarkNet architecture, with two classification heads added at the end. This multi-task proposal employs the Seeing Through Fog dataset[3] and estimates the weather and road conditions in 5 (clear, snow, light fog, dense fog, and rain) and 4 (dry, full snow, slushy, and wet) classes, respectively. The projected dataset comprises 12,997 representations with a resolution of 2048×64 pixels. Leveraging both the last and strongest LiDAR returns, the researchers reported a test accuracy of 76.71% for weather classification and 66.69% for road condition classification. Additionally, the model is stated to have 1.83 M parameters and runs at 102 FPS on a GeForce RTX 2070 graphics card.

In the study conducted by Dhananjaya et al. [14], an image dataset was acquired to estimate both weather conditions (fog, rain, and snow) and light levels (bright, moderate, and low). The data, captured in RCCC (red/clear) format, featured a resolution of 1024×1084 pixels. Additionally, an active learning framework was introduced for automated labeling and dataset redundancy reduction. This multi-task proposal relies on the ResNet18 pre-trained on the ImageNet dataset[4], taking grayscale-converted and resized images of 224×224 pixels as input. The authors reported test F_1 scores of 73%, 75%, and 78% for fog, rain, and snow estimation, respectively. Notably, there is no mention of the architecture's inference time.

In the work presented by Silva et al. [15], a convolutional neural network (CNN) named MobileWeatherNet was introduced. This CNN, consisting of 1.60 M parameters, is based on the architecture proposed by Simonyan and Zisserman [16]. The network employs the LiDAR bird's-eye view projection with dimensions $256 \times 256 \times 3$ to predict the weather conditions categorized into three classes: clear, fog, and rain. The authors reported a test accuracy of 97% and an inference time of 1.60 ms on the RADIATE dataset.[5] However, the specifications of the computer used for the time analysis are not provided.

Considering the well-balanced trade-off between accuracy performance and inference time presented by the RangeWeatherNet and MobileWeatherNet architectures, their respective encoders will serve as baselines for assessing the proposed architectures.

B. MOR Estimation

Currently, MOR estimation heavily relies on camera-based methods and neural networks, typically categorizing ranges in intervals of tens of meters [7]. In the study by Vaibhav et al. [2], the authors proposed a hybrid neural network for MOR estimation from camera images, addressing a 3-class classification task (0–50 m, 50–100 m, and >150 m). The network incorporates three types of input features: camera image, block-wise Shannon entropy, and block-wise discrete cosine transform. On their large-scale private dataset featuring highway and urban

[3] Seeing Through Fog Dataset. Available: https://www.uni-ulm.de/en/in/driveu/projects/dense-datasets/

[4] ImageNet Dataset. Available: https://www.image-net.org

[5] Heriot-Watt RADIATE Dataset. Available: https://pro.hw.ac.uk/radiate/

scenarios, the authors reported a test accuracy of 87.45% and an inference time of 1.06 seconds measured on an NVIDIA TITAN V graphics card, for an input image size of 1920 × 1080 pixels. Despite the adequate performance, the notably high inference time renders it impractical for real-time AD, motivating the current work.

In the work by Chaabani et al. [3], a deep learning model was introduced leveraging the AlexNet architecture for feature extraction and employing 5 multiclass one-vs-rest support vector machines (SVMs) for classification. Employing the publicly available FROSI synthetic dataset[6], the model is designed for MOR classification in 5 classes (<100 m, 100–200 m, 200–300 m, 300–400 m, >400 m). The dataset comprises 3,528 camera images of uncluttered road scenes with a resolution of 1400 × 600 pixels. The researchers reported a test accuracy of 99.02%; however, nothing was said about the inference time of the model. It is worth noting that this dataset lacks high photo-realism and data diversity.

C. *Attention for Autonomous Driving*

Attention mechanisms, inspired by the seminal work of Vaswani et al. [17], have gained significant attention and exploration in the field of AD. This interest is particularly evident in applications such as lane changing [18], object detection [19], [20], motion forecasting [21], [22], [23], [24], and waypoint prediction [25], [26]. In the study conducted by Prakash et al. [27], a novel architecture for end-to-end driving was introduced, consisting of two key elements: 1) a multi-modal fusion transformer entitled TransFuser, designed to integrate information from two modalities, namely single-view image and LiDAR Bird's Eye View (BEV), and 2) an auto-regressive network for waypoint prediction. TransFuser incorporates four transformer blocks with eight attention layers with four attention heads to fuse intermediate feature maps between both modalities. The fusion process is performed at four resolutions (64 × 64, 32 × 32, 16 × 16, and 8 × 8) throughout the ResNet feature extractor, resulting in a 512-dimensional feature vector output for each modality. These output vectors are then combined through element-wise summation. According to the authors, this approach led to remarkable performance, achieving excellent results on CARLA.[7] Precisely, the proposal yielded a driving score (DS) of 54.52% (±4.29%) and a route completion ratio (RC) of 78.41% (±3.75%) in Town05 Short. In the case of Town05 Long, the results were a DS of 33.15% (±4.04%) and an RC of 56.36% (±7.14%). Considering its versatility and effectiveness across various artificial intelligence (AI) tasks, the TransFuser encoder will be employed to evaluate the proposed architectures.

In the article by Choi et al. [28], the authors introduced the Cognitive TransFuser, a framework that builds upon the TransFuser architecture. As part of this framework, the researchers incorporated two auxiliary tasks, namely semantic segmentation and traffic light classification. According to the authors, these tasks were selected based on their semantic importance and strong correlation with the primary waypoint prediction task. In the proposed architecture, the semantic segmentation feature map is fused into the first transformer block, and the 512-dimensional feature vector output is utilized to predict both the traffic light label and local waypoints through a gated recurrent unit (GRU) sub-network. The researchers disclosed a DS of 80.67% and an RC of 95.14% in Town05 Short. In Town05 Long, the results were a 52.70% DS and a 96.18% RC. Additionally, the authors reported an inference time of 22.60 ms, measured on an NVIDIA GTX 1080Ti graphics card, utilizing an input data resolution of 256 × 256 for both modalities.

In their article, Shao et al. [29] introduced InterFuser, an architecture with ∼53 M parameters designed to process and fuse data from multi-modal multi-view sensors for comprehensive scene understanding and adversarial event detection. It comprises three main components: 1) a multi-modal multi-view fusion transformer encoder to fuse signals from multiple RGB cameras and LiDAR, 2) a transformer decoder to generate low-level actions and interpretable intermediate features, and 3) a safety controller to confine low-level control within the safe set, utilizing the interpretable intermediate features. In 2022, InterFuser achieved state-of-the-art performance on the Town05 benchmark, attaining a DS of 94.95% (±1.91%) and an RC of 95.19% (±2.57%) in Town05 Short, as well as 68.31% (±1.86%) and 94.97% (±2.87%) in Town05 Long, respectively. Additionally, the authors reported an inference time of approximately 20 ms on an NVIDIA GTX 1080Ti. Furthermore, in 2023, the same research group introduced ReasonNet [30], surpassing their previous effort by integrating temporal and global information from the driving scene. ReasonNet emerged as the leader in both the Town05 benchmark and the CARLA leaderboard (Sensors Track).

III. Methodology

A. *Proposed Architectures*

The proposed architectures are detailed in three parts: 1) the inputs, 2) the architectures per se, and 3) the outputs. For enhanced clarity, each part within the architecture diagrams is visually distinguished by a specific color: 1) blue for inputs, 2) green for architectures, and 3) orange for outputs.

1) Inputs: As illustrated in Figs. 1 and 2, the proposed architectures utilize the Shannon entropy derived from the camera image and the projected LiDAR intensity and range as input. After undergoing center cropping (0.5, 0.5) and normalization (0–1), the input data is fed directly into the proposed models. The architecture depicted in Fig. 1 employs early fusion via a concatenation process, enhancing rich information retention, and simplicity, and reducing computational complexity. Conversely, the architecture in Fig. 2 employs intermediate fusion with cross-modal self-attention, facilitating the fusion of the global context from both modalities to improve contextual understanding and optimize parameter usage. Center cropping is performed to expedite both the training and inference processes.

[6] FROSI Dataset. Available: https://www.livic.ifsttar.fr/linstitut/cosys/laboratoires/livic-ifsttar/logiciels/bases-de-donnees/frosi/

[7] CARLA Autonomous Driving Leaderboard. Available: https://leaderboard.carla.org

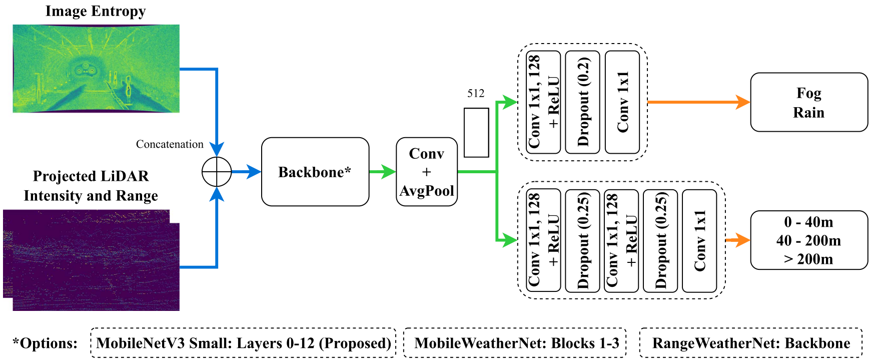

Fig. 1. Illustration of the multi-task architectures featuring early fusion. The proposed architecture adopts the MobileNetV3-Small encoder as backbone, while alternative backbone options, namely the MobileWeatherNet and RangeWeatherNet encoders, are employed as baselines.

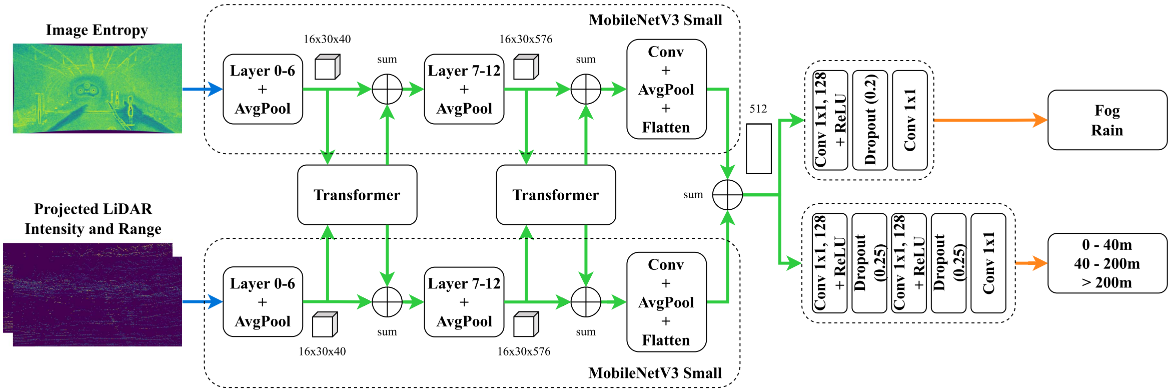

Fig. 2. Illustration of the proposed multi-task architecture featuring intermediate fusion. It uses the MobileNetV3-Small encoder and cross-modal self-attention mechanisms for feature fusion.

The Shannon entropy is intended to highlight regions of the image with uniform pixel values (low entropy) and regions with variable and complex pixel patterns (high entropy). The dimension of the transformation output is the same as that of the original image. Fog reduces contrast and homogenizes pixel values, resulting in lower entropy values in foggy regions of the camera image. Rain, on the other hand, introduces localized variations in intensity and texture, increasing entropy in the regions of the image affected by raindrops. It is noteworthy to state that in the real world, fog and rain can occur together to varying degrees and that several factors, such as lighting conditions and camera settings, can influence the image's appearance.

Regarding the LiDAR data, it is projected into the camera coordinates with the missing measurements encoded with zero value. This strategy was implemented given that the bird's eye-view projection or the raw point-cloud representation do not easily allow for deep early fusion as they are inherently different from the camera features. The dimensions of the LiDAR range view projections are also the same as that of the original image. Both fog and rain can scatter and absorb the laser pulses resulting in decreased intensity readings and distorted point clouds. However, in some cases, the effects due to rain can be greater, mainly due to the multiple reflections caused by raindrops on surfaces.

2) Architectures: Two architectures are proposed, differentiating themselves through variations in backbone design and multi-modal fusion techniques. One favors optimal performance with minimal resource usage, while the other aims for maximum performance without excessive consumption of computing resources.

Fundamentally, the simplest architecture consists of a MobileNetV3-Small (layers 0-12) as backbone and two task-specific heads, see Fig. 1. The heads are designed by taking a 512-dimensional vector as input and employing a combination of convolution and dropout layers. The selected non-linear activation function is the Rectified Linear Unit (ReLU). As a reference, Google's MobileNetV3 was presented at ICCV in Seoul, South Korea in 2019 [31]. MobileNetV3-Small is a widely adopted network designed to be executed on embedded systems with limited resources while maintaining competitive performance in tasks such as image classification and object detection.

Conversely, the more complex architecture, inspired by the work of Prakash et al. [27], integrates two dedicated MobileNetV3-Small feature extractors, one for each modality, and performs cross-modal self-attention intermediate fusion at two distinct resolutions, as illustrated in Fig. 2. Additionally, it employs the same two task-specific heads. Each transformer

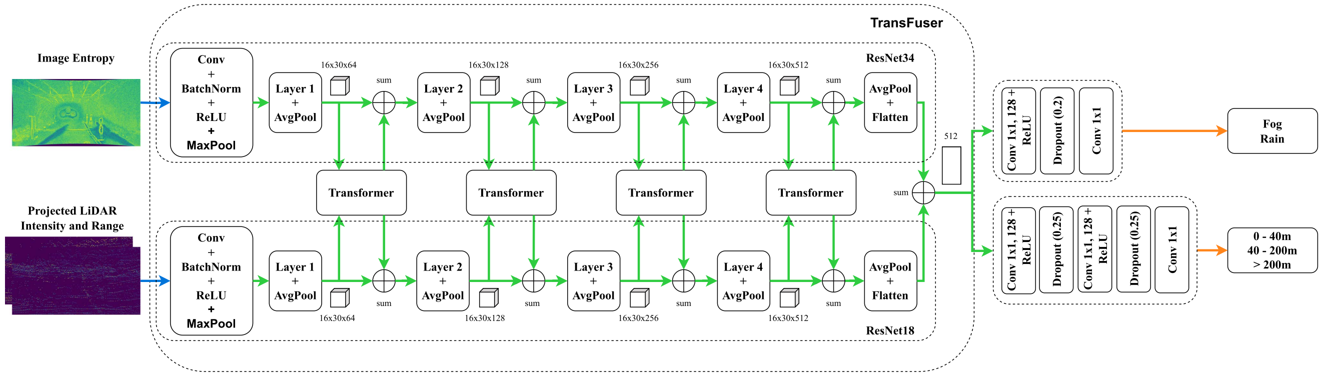

Fig. 3. Illustration of the baseline architecture featuring the TransFuser encoder.

block comprises four attention layers with two attention heads. Following the methodology presented in [27], the higher resolution feature maps from the early encoder blocks are subjected to downsampling via average pooling to a fixed resolution (16 × 30) before passing on as inputs to the transformer block. The output from the transformer block is then upsampled to the original resolution using bilinear interpolation, followed by an element-wise summation with the pre-existing feature maps. After feature fusion at various resolutions and the subsequent flattening operation applied to the feature maps of each modality, a 512-dimensional feature vector is generated via element-wise summation. This process results in a compact representation of the vehicle's surroundings, which serves as input for the two classification heads.

3) Outputs: As previously stated, weather estimation is tackled as a binary classification task, distinguishing between fog and rain. Meanwhile, the MOR estimation entails a multi-class classification with three ordinal classes (0-40 m, 40-200 m, and >200 m). Therefore, the weather classification head comprises 2 outputs, while the MOR classification head comprises 3 outputs.

Considering the influence of the TransFuser architecture in this study, coupled with its notable performance across diverse AD tasks, a direct comparison is conducted with an architecture featuring the TransFuser encoder along with the two classification heads previously mentioned, as depicted in Fig. 3. Furthermore, the proposed architectures are compared with adapted versions of the MobileWeatherNet and RangeWeatherNet architectures documented in the literature, as showcased in Fig. 1. These architectures serve as a baseline for evaluating the performance of the proposed architectures. Additionally, a direct comparison is made with the equivalent single-task architectures to assess the advantages and disadvantages associated with adopting multi-task architectures. In the multi-task proposals, the backbone is shared between the two classification heads, whereas in the single-task architectures, there are two separate backbones, one for each classification head.

B. Attentive Multi-Modal Fusion Transformer

The transformer architecture is designed to process an input sequence consisting of discrete tokens (data patches), each representing a feature vector. To capture cross-token spatial dependencies and account for the sequential order of tokens, a learnable positional encoding is introduced by performing element-wise summation with the input embeddings. In formal terms, the input sequence is denoted as $F^{in} \in \mathbb{R}^{N \times D_f}$, where N is the number of tokens, and each token is represented by a feature vector of dimensionality D_f. Linear projections are used to compute a set of queries, keys, and values (Q, K, and V). In addition, the transformer employs scaled dot products between Q and K to compute the attention weights and aggregates the values for each query,

$$A = \text{softmax}\left(\frac{QK^T}{\sqrt{d}}\right)V \tag{1}$$

where Q, K, V denote the query, key, and value matrices and d the dimensions of the query/key.

Subsequently, a non-linear transformation is applied to produce the output features, F^{out}, maintaining the same shape as that of F^{in}:

$$F^{out} = \text{MLP}(A) + F^{in} \tag{2}$$

The attention mechanism is repeated multiple times in the architecture, leading to a total of L attention layers. Within each layer, multiple parallel attention heads are employed, generating several sets of Q, K, and V values for each F^{in}. Fig. 4 offers a comprehensive depiction of the entire fusion process at a single scale.

C. Losses

As presented in the previous sub-section, the architectures presented have two outputs: weather and MOR. For the weather, the Focal Loss (FL) is employed during the training process to address the challenges posed by the class imbalance in the dataset. Concerning the MOR, the training is conducted using the Ordinal Encoding Loss (OEL).

The FL is a modification of the standard cross entropy loss that aims to address the problem of class imbalance in classification tasks. It assigns higher weights to hard-to-classify examples and reduces the impact of well-classified examples. The FL,

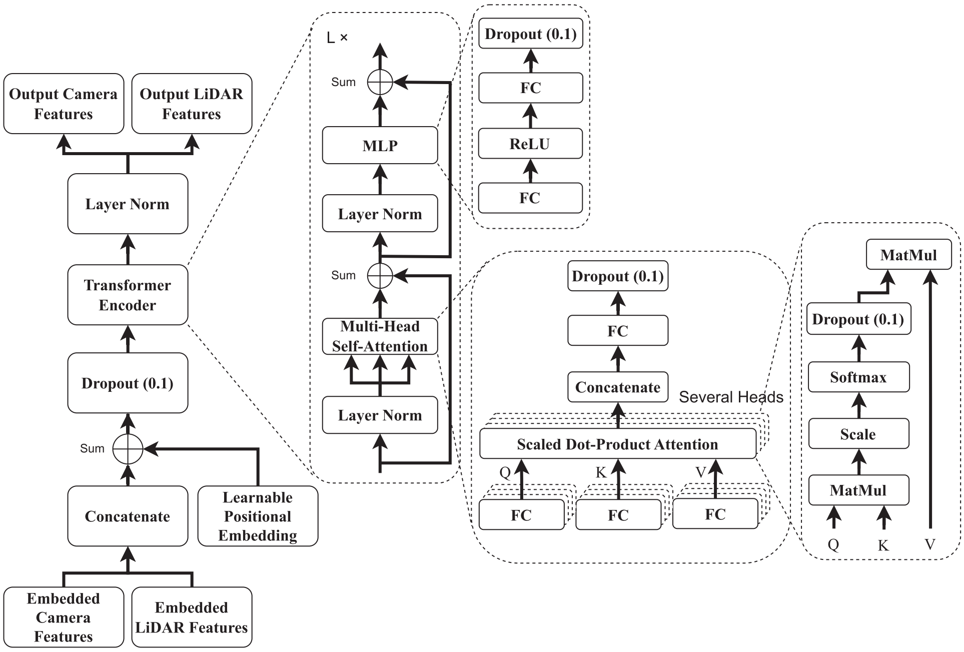

Fig. 4. Detailed illustration of the transformer block.

presented in (3), was first introduced in Lin et al. [32].

$$\mathcal{L}_F(y_k, p_k) = -\sum_{k=1}^{K} [\alpha_k \cdot (1 - p_k)^{\gamma} \cdot y_k \log(p_k))] \quad (3)$$

where y_k is the k-th element of the one-hot encoded true label (that is, $y_k = 1$ if k is the true class and $y_k = 0$ otherwise), p_k is the k-th element of the predicted probability distribution, K is the number of classes, α_k is the class-dependent scaling factor that adjusts the balance between easy and hard examples and γ (≥ 0) is the focusing parameter that smoothly adjusts the rate at which easy examples are down-weighted.

The OEL is employed to encourage ordinality in the prediction distributions generated by the network [33]. For instance, predicting that both "0-40 m" and ">200 m" are more probable than "40-200 m" would not be logically consistent. To this end, the network is trained to produce $K-1$ outputs, each of which produces a binary decision between adjacent classes. In ordinal encoding, classes are encoded using a cumulative distribution approach: if k^* is the true class, then $y_k = 1$ if $k < k^*$ and $y_k = 0$ otherwise. Each output of the network represents the incremental neighbor probability. The inverse operation, conducted during inference to predict the true class $\hat{k}$, involves summing these outputs, defined as $\hat{k} = \sum_{k=1}^{K-1} \mathbb{1}(p_k \geq 0.5)$. The multi-class Cross Entropy loss is then used to optimize the neural network.

The multi-task training is conducted by simultaneously optimizing the Focal and Ordinal Encoding losses. Adaptive gradient-based optimization algorithms, like AdamW, dynamically adjust the learning rate on a per-parameter basis, effectively reducing the learning rate for frequently updated parameters and increasing it for less frequently updated ones. This is achieved by scaling the global learning rate by a function of the past gradients specific to each parameter. As indicated by to A´ lvaro S. Hervella et al. [34], the gradients of both tasks can be decoupled to leverage this normalization for multi-task learning. This decoupling allows the calculation of task-specific per-parameter learning rates, taking into account only the past gradients associated with each specific task. According to the authors, this multi-adaptive (M-Ada) technique ensures balanced training without the need for additional hyperparameters. Consequently, the global parameters are updated as:

$$\theta_i^{t+1} = \theta_i^t + \Delta\theta_{i,W}^t \left(\eta_{\theta_{i,W}}^t\right) + \Delta\theta_{i,MOR}^t \left(\eta_{\theta_{i,MOR}}^t\right) \quad (4)$$

where $\Delta\theta_{i,W}^t$ and $\Delta\theta_{i,MOR}^t$ denote the parameters updates due to the weather and MOR estimation tasks, respectively.

However, as highlighted in [34], this M-Ada optimization technique requires additional training memory, as well as additional operations to compute the task-specific gradients and parameter updates. As a result, the computational time required for each training experiment is prolonged compared to using fixed loss weighting (FLW) hyperparameters. To evaluate the effectiveness of the M-Ada optimization technique, a direct comparison is conducted against the conventional FLW optimization technique.

IV. Experiments

A. Dataset

Our proposed architectures are explored on the publicly available FogChamber dataset [8] for weather and MOR classification. The dataset was acquired with an OnSemi AR0230 camera and a Velodyne HDL64E S3D LiDAR in a controlled weather chamber

[8] DENSE Datasets. Available: https://www.uni-ulm.de/en/in/driveu/projects/dense-datasets

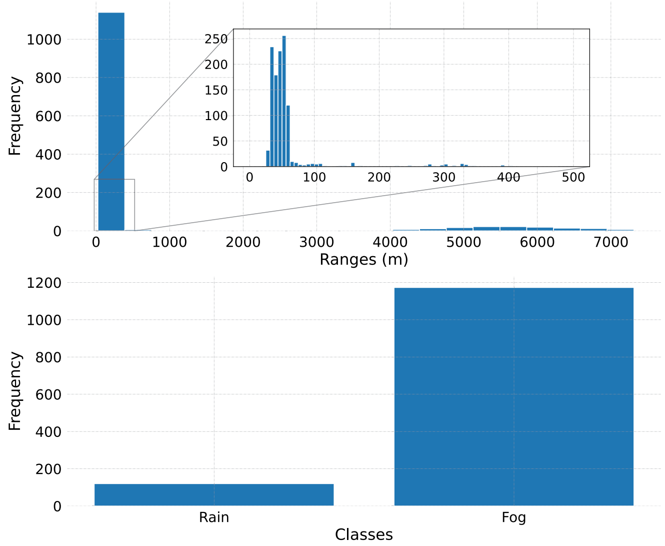

Fig. 5. Histograms depicting MOR data (top) and weather data (bottom). The top histogram provides a closer view within the 0-500 m range.

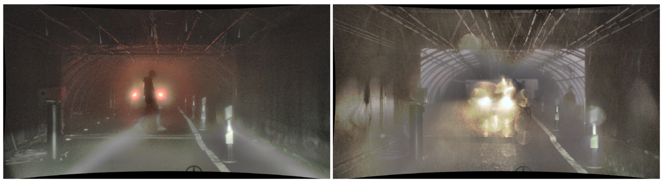

Fig. 6. Images captured by the camera under different weather conditions: (left) a sample with fog and (right) a sample with rain. The MOR for the rain sample is 56 meters, while for the fog sample, it is 34 meters.

TABLE I
SENSOR SPECIFICATIONS USED IN THE FOGCHAMBER DATASET

OnSemi AR0230 (with Lensagon B5M8018C optics)	
Field of view	39.6°×21.7°
Focal length	8 mm
Frequency	30 Hz
Quantization	12 bit
Resolution	1920x1024 pixels
Velodyne HDL64E S3D	
Angular resolution	0.4°
Channels	64
Frequency	10 Hz
Horizontal field of view	360°
Measurement range	120 m
Returns	Dual (strongest and last)
Vertical field of view	26.9° (2.0° to -24.9°)
Wavelength	903 nm

TABLE II
SUMMARY OF THE EMPLOYED HYPERPARAMETERS

Hyperparameters	
Input Shape	$1024 \times 1920 \times 3$
Model Input Shape	$512 \times 960 \times 3$
Losses	Focal / Ordinal Encoding
Optimizers	AdamW / AdamW
Global Learning Rates	1e−5 / 1e−5
Weight Decays	1e−4 / 1e−4
Batch Size	16
Number of Epochs	50
Loss Weighting Hyperparam.	$\gamma_{Weather} = 1$, $\gamma_{MOR} = 1$
Seeds	10, 21, 29, 38, 64, 65, 70, 72, 81, 89

TABLE III
PARAMETER COUNT FOR EACH ARCHITECTURE

Task	Fusion	Encoder	N. of Parameters
S	E	MobileWeatherNet **[15]**	1,150,802
S	*E*	*MobileNetV3-Small (Proposed)*	2,593,220
S	E	RangeWeatherNet **[13]**	7,862,884
S	*I*	*MobileNetV3-Small (Proposed)*	27,612,676
S	I	TransFuser **[27]**	133,945,348
M	E	MobileWeatherNet **[15]**	649,579
M	*E*	*MobileNetV3-Small (Proposed)*	1,370,788
M	E	RangeWeatherNet **[13]**	4,005,620
M	*I*	*MobileNetV3-Small (Proposed)*	13,880,516
M	I	TransFuser **[27]**	67,046,852

Notes: Italics denote the proposed architectures; S – single-task; M – multi-task; E – early fusion; I – intermediate fusion.

under varying weather and lighting conditions. A sample of the dataset is depicted in Fig. 6 with the different weather and lighting conditions (day and night). Details on the weather chamber setup can be found in Colomb et al. [35] and Duthon et al. [36]. The sensor specifications are listed in Table I. The dataset has 1293 samples with two weather labels (fog and rain) and a median MOR of 49 meters; however, as shown in Fig. 5, the weather classes are highly imbalanced. The dataset imbalance can negatively affect the model's performance due to the model's inherent bias towards the majority class [37]. To mitigate this problem, the minority class was oversampled (as described in the next subsection). To the best of our knowledge, the FogChamber dataset is the only publicly available dataset that features camera and LiDAR data, along with weather and MOR annotations.

B. Implementation

The FogChamber dataset was randomly split into three sets, with 60% allocated for training, 20% for validation, and 20% for testing. The training set was subjected to a random synthetic augmentation, which consisted of applying horizontal flips and affine transformations, with a scaling factor ranging between 1.10 and 1.25. To mitigate the problem of imbalanced weather classes, a weighted sampling technique was employed. This technique assigns higher weights to samples from the minority class (i.e., rain), with the weights being calculated inversely proportional to the class frequency. This oversampling approach results in a more balanced class distribution within the training set. Furthermore, all input data were center-cropped (0.5, 0.5) and normalized (0–1) based on the statistics from the training set.

All experiments were trained from scratch with the same set of hyperparameters across ten different seeds. A summary of the employed hyperparameters is provided in Table II. Furthermore, the FL was employed with $\alpha_k = 1$ and $\gamma = 2$. Both losses were optimized using the AdamW algorithm with $\text{ß}_1 = 0.9$ and $\text{ß}_2 = 0.999$. The top-performing models were saved based on the validation set losses.

As a reference, the parameter count for both multi-task and single-task variants of each architecture is presented in Table III.

TABLE IV
TEST RESULTS FOR WEATHER AND MOR CLASSIFICATION AND COMPARISON WITH LITERATURE SINGLE-TASK AND MULTI-TASK ALTERNATIVES

Task	Fusion	Encoder	Optimization	Weather			MOR			Memory (MB)	Time (ms)
				Accuracy	Cohen Kappa	F1	Accuracy	Cohen Kappa	F1		
S	E	MobileWeatherNet **[15]**	N/A	94.58% (±4.34%)	73.67% (±13.87%)	95.04% (±3.48%)	79.04% (±3.86%)	62.12% (±9.12%)	78.75% (±5.60%)	**4.42 (±0.00)**	6.67 (±0.01)
S	E	RangeWeatherNet **[13]**	N/A	89.65% (±5.99%)	20.38% (±32.70%)	88.03% (±5.57%)	69.00% (±4.42%)	29.33% (±9.40%)	62.89% (±4.78%)	31.03 (±0.00)	18.25 (±0.02)
S	*E*	*MobileNetV3-Small (Proposed)*	*N/A*	**98.81% (±0.76%)**	**92.77% (±4.17%)**	**98.82% (±0.73%)**	**90.38% (±1.62%)**	**81.93% (±3.48%)**	**90.35% (±1.68%)**	10.10 (±0.00)	**4.80 (±0.05)**
S	I	TransFuser **[27]**	N/A	**99.85% (±0.19%)**	**99.01% (±1.25%)**	**99.84% (±0.19%)**	**92.58% (±1.10%)**	**86.03% (±2.42%)**	**92.58% (±1.09%)**	515.53 (±0.00)	130.66 (±0.05)
S	*I*	*MobileNetV3-Small (Proposed)*	*N/A*	99.50% (±0.42%)	96.76% (±3.16%)	99.49% (±0.43%)	91.38% (±1.25%)	83.60% (±2.93%)	91.30% (±1.30%)	**106.92 (±0.00)**	**31.36 (±0.07)**
M	E	MobileWeatherNet **[15]**	M-Ada	93.85% (±2.19%)	70.93% (±7.53%)	94.45% (±1.91%)	79.85% (±5.27%)	65.20% (±8.46%)	80.61% (±4.79%)	**2.50 (±0.00)**	3.37 (±0.00)
M	E	MobileWeatherNet **[15]**	FLW	90.38% (±3.82%)	59.98% (±9.34%)	91.71% (±2.93%)	83.19% (±5.40%)	70.25% (±9.02%)	83.64% (±5.03%)	**2.50 (±0.00)**	3.33 (±0.02)
M	E	RangeWeatherNet **[13]**	M-Ada	81.42% (±4.16%)	18.14% (±12.42%)	83.64% (±2.84%)	66.69% (±2.72%)	33.44% (±8.47%)	63.73% (±4.44%)	16.24 (±0.00)	9.06 (±0.07)
M	E	RangeWeatherNet **[13]**	FLW	77.23% (±5.47%)	9.01% (±8.81%)	80.48% (±3.64%)	69.00% (±3.49%)	37.31% (±6.47%)	67.25% (±2.80%)	16.24 (±0.00)	9.09 (±0.02)
M	*E*	*MobileNetV3-Small (Proposed)*	*M-Ada*	**98.88% (±1.05%)**	**94.02% (±4.72%)**	**98.92% (±0.98%)**	89.77% (±1.14%)	80.83% (±2.15%)	89.76% (±1.19%)	5.33 (±0.00)	**2.50 (±0.02)**
M	*E*	*MobileNetV3-Small (Proposed)*	*FLW*	96.92% (±1.34%)	83.90% (±5.54%)	97.09% (±1.21%)	**91.15% (±1.49%)**	**83.43% (±2.48%)**	**91.14% (±1.45%)**	5.33 (±0.00)	**2.50 (±0.02)**
M	I	TransFuser **[27]**	M-Ada	99.19% (±0.99%)	95.66% (±4.87%)	99.22% (±0.93%)	**92.08% (±0.91%)**	**85.06% (±1.82%)**	**92.06% (±0.92%)**	258.30 (±0.00)	65.32 (±0.06)
M	I	TransFuser **[27]**	FLW	**99.73% (±0.39%)**	**98.57% (±2.09%)**	**99.73% (±0.38%)**	91.65% (±1.71%)	84.43% (±3.30%)	91.71% (±1.60%)	258.30 (±0.00)	65.36 (±0.05)
M	*I*	*MobileNetV3-Small (Proposed)*	*M-Ada*	99.38% (±0.39%)	96.24% (±2.47%)	99.38% (±0.39%)	91.88% (±1.01%)	84.75% (±2.06%)	91.88% (±0.95%)	**54.06 (±0.00)**	**15.55 (±0.19)**
M	*I*	*MobileNetV3-Small (Proposed)*	*FLW*	97.00% (±0.75%)	83.68% (±4.80%)	97.17% (±0.69%)	91.88% (±1.21%)	84.66% (±2.39%)	91.84% (±1.22%)	**54.06 (±0.00)**	**15.60 (±0.05)**

Notes: Italics denote the proposed architectures and bold denote the best results for each category; S – single-task; M – multi-task; E – early fusion; I – intermediate fusion.

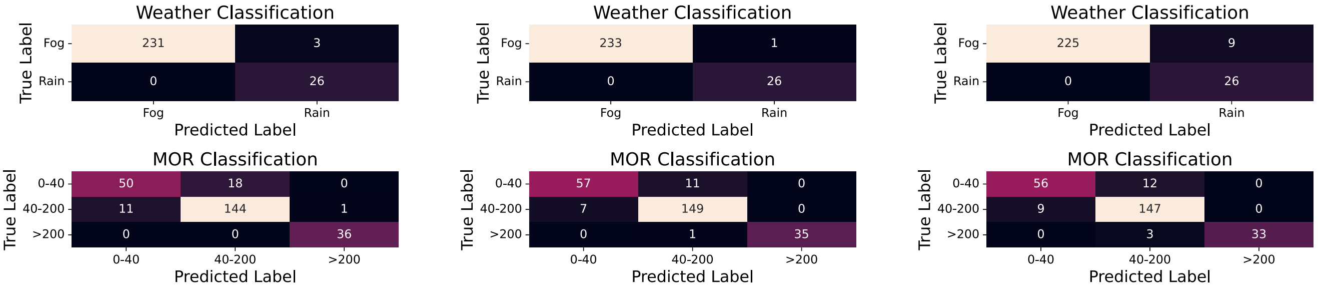

Fig. 7. Confusion matrices of the multi-task architectures featuring the M-Ada optimization (seed 10). From left to right: (left) architecture with the MobileNetV3-Small encoder and early fusion, (center) architecture with the MobileNetV3-Small encoder and cross-modal self-attention intermediate fusion, and (right) architecture with the TransFuser encoder.

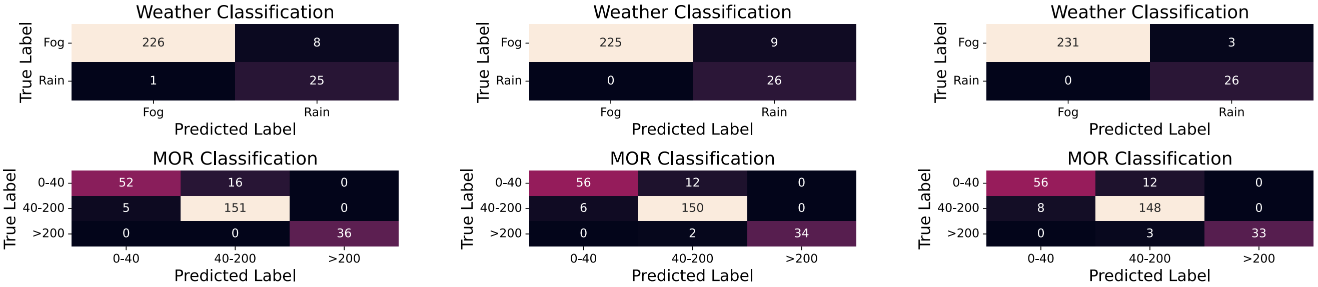

Fig. 8. Confusion matrices of the multi-task architectures featuring the FLW optimization (seed 10). From left to right: (left) architecture with the MobileNetV3-Small encoder and early fusion, (center) architecture with the MobileNetV3-Small encoder and cross-modal self-attention intermediate fusion, and (right) architecture with the TransFuser encoder.

C. Results

The test results from the different experiments are presented in Table IV, featuring mean and standard deviation values, with the best results highlighted in bold. The evaluation of the models involves the use of metrics such as weighted accuracy, weighted F_1, and Cohen's kappa (κ) score. These metrics are commonly adopted in classification scenarios involving imbalanced datasets. The memory allocation, measured in megabytes, and the average inference time, measured in milliseconds, are also provided. It is noteworthy to state that the average inference time was calculated based on 10,000 repetitions with GPU warm-up. Moreover, the confusion matrices for the three best-performing encoders can be seen in Figs. 7– 9.

Among the results obtained, it is worth emphasizing the following:

- The proposed multi-task architectures outperform the equivalent single-task architectures regarding inference time and memory allocation while maintaining comparable accuracy performance. Improvements in memory allocation range from 47.23% to 49.44%, and in average inference time from 47.92% to 50.41%, progressing from the simplest to the most complex architecture. This improvement can be attributed to the smaller number of parameters in the multi-task architectures;
- Among the early fusion architectures, the proposed approach based on the MobileNetV3-Small encoder offered the best results regarding both accuracy performance and inference times, while also presenting competitive memory allocation results;
- Despite exhibiting a slight decrease in accuracy performance compared to the alternative proposed architecture, the proposed early fusion architecture is significantly lighter, making it a potentially advantageous compromise in various applications;

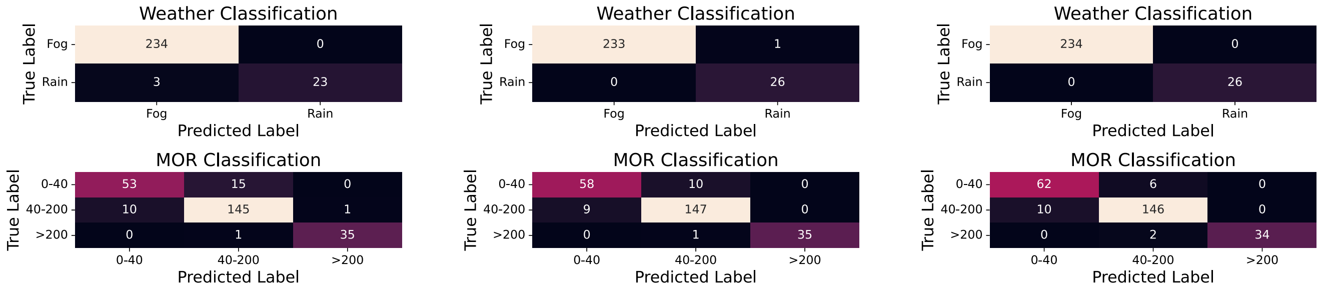

Fig. 9. Confusion matrices of the equivalent single-task architectures (seed 10). From left to right: (left) architecture with the MobileNetV3-Small encoder and early fusion, (center) architecture with the MobileNetV3-Small encoder and cross-modal self-attention intermediate fusion, and (right) architecture with the TransFuser encoder.

- The proposed early fusion architecture achieves an average inference time of 2.50 ms, equivalent to approximately 400 FPS, making it the fastest architecture assessed. Given the LiDAR acquisition rate of 10 FPS, achieving real-time inference is feasible even on a less powerful computer system;
- The proposed intermediate fusion architecture provides performance on par with the more complex baseline architecture incorporating the TransFuser encoder. Additionally, it holds the advantage of demanding significantly fewer resources while quadrupling the refresh rate;
- The proposed architectures show state-of-the-art results in both tasks. These strong performances are further evidenced by the clearly defined diagonals in the confusion matrices;
- With this particular set of hyperparameters, the baseline architecture incorporating the RangeWeatherNet encoder exhibits overfitting to the training set. This phenomenon is likely attributed to its high parameter count, leading to low Cohen's kappa scores;
- The baseline architecture featuring the MobileWeatherNet encoder requires the least memory among the assessed architectures; however, it exhibits longer inference times compared to the proposed early fusion architecture, and achieves significantly lower levels of performance;
- In specific experiments, the M-Ada optimization technique exhibits superior performance, while in others, the FLW optimization technique outperforms. Consequently, there is no significant improvement in performance that justifies the increased complexity and memory allocation required by the M-Ada optimization technique during the training phase.

When estimating these tasks, enhancing the complexity of the architectures leads to an improvement in accuracy performance. However, this enhancement is relatively modest when compared to the simultaneous substantial increases in memory allocation and inference time. For instance, when comparing the proposed early fusion architecture to the baseline featuring the TransFuser encoder, there is a 0.31% improvement in weather estimation and a 2.57% improvement in MOR estimation. Nonetheless, this improvement is offset by a striking 2,513% increase in inference time. Such a significant increase in resource consumption diminishes the practical value of the architectures in AD and may ultimately render them impractical.

TABLE V
SYSTEM SPECIFICATIONS EMPLOYED FOR TRAINING, VALIDATION, AND TESTING PURPOSES

System Specifications	
CPU	Intel Core i7-12700
RAM (GB)	16
GPU	NVIDIA RTX A2000
VRAM (GB)	12
PyTorch Version	1.13.1
CUDA Version	11.8
Python Version	3.10.12

The system specifications employed in this article are detailed in Table V.

V. CONCLUSION

In AD, algorithms must deliver peak performance while maximizing efficiency. Based on this consideration, this article introduces two deep learning-based architectures, employing early and intermediate sensor fusion and multi-task strategies, designed for concurrent weather and MOR classification. To the best of our knowledge, these architectures stand as the pioneering endeavor in tackling both of these tasks simultaneously in the context of AD. These architectures distinguish themselves through variations in backbone design and multimodal fusion techniques. One prioritizes achieving optimum performance with minimal resource usage, while the other aims for peak performance without excessive computational resource consumption.

The proposed architectures exhibit state-of-the-art results in both tasks and outperform the equivalent single-task architectures in inference time and memory allocation while maintaining comparable accuracy performance. Despite exhibiting slightly lower accuracy performance, the proposed early fusion architecture is ten times lighter and six times faster than its more complex counterpart, presenting a potentially advantageous compromise across various applications. Conversely, the proposed intermediate fusion architecture matches the performance of the baseline architecture featuring the literature TransFuser encoder. Furthermore, it holds the advantage of demanding significantly fewer resources while quadrupling the refresh rate.

Overall, the proposed architectures strike a commendable balance between accuracy performance, inference time, and memory allocation, making them highly suitable for AD. Specifically, the architecture incorporating the MobileNetV3-Small and early fusion achieves a noteworthy 98.88% accuracy in weather classification and 89.77% in MOR classification, with a competitive memory allocation of 5.33 MB and an inference time of 2.50 ms (400 FPS). Conversely, the architecture featuring the MobileNetV3-Small and cross-modal self-attention intermediate fusion demonstrates an even higher accuracy of 99.38% in weather classification and 91.88% in MOR classification, albeit with a higher memory allocation of 54.06 MB and an inference time of 15.55 ms (64 FPS).

VI. Future Research Directions

For future research endeavors, it would be interesting to conduct an analysis utilizing supplementary datasets acquired in real-world scenarios and covering a broader spectrum of weather conditions. Using an appropriate dataset, estimating weather across additional classes, specifically clear and snow would be compelling. Furthermore, it would be interesting to analyze the trade-offs between accuracy performance, inference time, and memory allocation by systematically reducing the layers within the transformers.

References

[1] I. Gultepe et al., “A review of high impact weather for aviation meteorology,” *Pure Appl. Geophys.*, vol. 176, pp. 1869–1921, 2019.

[2] V. Vaibhav, K. R. Konda, C. Kondapalli, K. Praveen, and B. Kondoju, “Real-time fog visibility range estimation for autonomous driving applications,” in *Proc. IEEE 23rd Int. Conf. Intell. Transp. Syst.*, 2020, pp. 1–6.

[3] H. Chaabani, N. Werghi, F. Kamoun, B. Taha, F. Outay, and A.-U.-H. Yasar, “Estimating meteorological visibility range under foggy weather conditions: A deep learning approach,” *Procedia Comput. Sci.*, vol. 141, pp. 478–483, 2018.

[4] L. Wang et al., “Multi-modal 3D object detection in autonomous driving: A survey and taxonomy,” *IEEE Trans. Intell. Veh.*, vol. 8, no. 7, pp. 3781–3798, Jul. 2023.

[5] J. Kocić, N. Jovičić, and V. Drndarević, “Sensors and sensor fusion in autonomous vehicles,” in *Proc. IEEE 26th Telecommun. Forum*, 2018, pp. 420–425.

[6] L. R. Agostinho, N. M. Ricardo, M. I. Pereira, A. Hiolle, and A. M. Pinto, “A practical survey on visual odometry for autonomous driving in challenging scenarios and conditions,” *IEEE Access*, vol. 10, pp. 72182–72205, 2022.

[7] Y. Zhang, A. Carballo, H. Yang, and K. Takeda, “Perception and sensing for autonomous vehicles under adverse weather conditions: A survey,” *ISPRS J. Photogrammetry Remote Sens.*, vol. 196, pp. 146–177, 2023.

[8] F. Sezgin, D. Vriesman, D. Steinhauser, R. Lugner, and T. Brandmeier, “Safe autonomous driving in adverse weather: Sensor evaluation and performance monitoring,” in *Proc. IEEE Intell. Veh. Symp.*, 2023, pp. 1–6.

[9] R. Heinzler, P. Schindler, J. Seekircher, W. Ritter, and W. Stork, “Weather influence and classification with automotive lidar sensors,” in *Proc. IEEE Intell. Veh. Symp.*, 2019, pp. 1527–1534.

[10] M. Jokela, M. Kutila, and P. Pyykönen, “Testing and validation of automotive point-cloud sensors in adverse weather conditions,” *Appl. Sci.*, vol. 9, no. 11, 2019, Art. no. 2341.

[11] M. Bijelic, T. Gruber, and W. Ritter, “A benchmark for lidar sensors in fog: Is detection breaking down?,” in *Proc. IEEE Intell. Veh. Symp.*, 2018, pp. 760–767.

[12] M. Kutila, P. Pyykönen, H. Holzhüter, M. Colomb, and P. Duthon, “Automotive LiDAR performance verification in fog and rain,” in *Proc. IEEE 21st Int. Conf. Intell. Transp. Syst.*, 2018, pp. 1695–1701.

[13] G. Sebastian, T. Vattem, L. Lukic, C. Bürgy, and T. Schumann, “RangeWeatherNet for LiDAR-only weather and road condition classification,” in *Proc. IEEE Intell. Veh. Symp.*, 2021, pp. 777–784.

[14] M. M. Dhananjaya, V. R. Kumar, and S. K. Yogamani, “Weather and light level classification for autonomous driving: Dataset, baseline and active learning,” in *Proc. IEEE Int. Intell. Transp. Syst. Conf.*, 2021, pp. 2816–2821.

[15] M. P. Silva, D. Carneiro, J. Fernandes, and L. F. Texeira, “MobileWeatherNet for lidar-only weather estimation,” in *Proc. IEEE Int. Joint Conf. Neural Netw.*, 2023, pp. 1–8.

[16] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” in *Proc. 3rd Int. Conf. Learn. Representations*, 2015, pp. 1–14.

[17] A. Vaswani et al., “Attention is all you need,” in *Proc. 31st Int. Conf. Neural Inf. Process. Syst.*, Long Beach, CA, USA, 2017, pp. 6000–6010.

[18] Y. Chen, C. Dong, P. Palanisamy, P. Mudalige, K. Muelling, and J. M. Dolan, “Attention-based hierarchical deep reinforcement learning for lane change behaviors in autonomous driving,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops*, 2019, pp. 1326–1334.

[19] S. Chen, S. Zhang, J. Shang, B. Chen, and N. Zheng, “Brain-inspired cognitive model with attention for self-driving cars,” *IEEE Trans. Cogn. Devlop. Syst.*, vol. 11, no. 1, pp. 13–25, Mar. 2019.

[20] L. L. Li et al., “End-to-end contextual perception and prediction with interaction transformer,” in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2020, pp. 5784–5791.

[21] A. Sadeghian, V. Kosaraju, A. Sadeghian, N. Hirose, S. H. Rezatofighi, and S. Savarese, “SoPhie: An attentive GAN for predicting paths compliant to social and physical constraints,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 1349–1358.

[22] A. Sadeghian, F. Legros, M. Voisin, R. Vesel, A. Alahi, and S. Savarese, “CAR-Net: Clairvoyant attentive recurrent network,” in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 151–167.

[23] Y. Huang, H. Bi, Z. Li, T. Mao, and Z. Wang, “STGAT: Modeling spatial-temporal interactions for human trajectory prediction,” in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2019, pp. 6271–6280.

[24] C. Choi and B. Dariush, “Looking to relations for future trajectory forecast,” in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2019, pp. 921–930.

[25] K. Chitta, A. Prakash, B. Jaeger, Z. Yu, K. Renz, and A. Geiger, “TransFuser: Imitation with transformer-based sensor fusion for autonomous driving,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 11, pp. 12878–12895, Nov. 2023.

[26] J. Du, Y. Zhao, and H. Cheng, “Target-point attention transformer: A novel trajectory predict network for end-to-end autonomous driving,” 2023, *arXiv:2308.01496*.

[27] A. Prakash, K. Chitta, and A. Geiger, “Multi-modal fusion transformer for end-to-end autonomous driving,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 7077–7087.

[28] H.-S. Choi, J. Jeong, Y. H. Cho, K.-J. Yoon, and J.-H. Kim, “Cognitive TransFuser: Semantics-guided transformer-based sensor fusion for improved waypoint prediction,” 2023, *arXiv:2308.02126*.

[29] H. Shao, L. Wang, R. Chen, H. Li, and Y. Liu, “Safety-enhanced autonomous driving using interpretable sensor fusion transformer,” in *Proc. 6th Conf. Robot Learn.*, 2022, pp. 726–737.

[30] H. Shao, L. Wang, R. Chen, S. L. Waslander, H. Li, and Y. Liu, “ReasonNet: End-to-end driving with temporal and global reasoning,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2023, pp. 13723–13733.

[31] A. Howard et al., “Searching for MobileNetV3,” in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2019, pp. 1314–1324.

[32] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, “Focal loss for dense object detection,” in *Proc. IEEE Int. Conf. Comput. Vis.*, 2018, pp. 2980–2988.

[33] T. Albuquerque, R. Cruz, and J. Cardoso, “Ordinal losses for classification of cervical cancer risk,” *PeerJ Comput. Sci.*, vol. 7, 2021, Art. no. e457.

[34] A. S. Hervella, J. Rouco, J. Novo, and M. Ortega, “End-to-end multi-task learning for simultaneous optic disc and cup segmentation and glaucoma classification in eye fundus images,” *Appl. Soft Comput.*, vol. 116, 2022, Art. no. 108347.

[35] M. Colomb, J. Dufour, M. Hirech, P. Lacôte, P. Morange, and J. Boreux, “An innovative artificial fog production device improved in the European project fog,” *Atmos. Res.*, no. 87, pp. 242–251, Aug. 2008, doi: 10.1016/j.atmosres.2007.11.021.

[36] P. Duthon, F. Bernardin, F. Chausse, and M. Colomb, “Methodology used to evaluate computer vision algorithms in adverse weather conditions,” *Transp. Res. Procedia*, vol. 14, pp. 2178–2187, 2016.

[37] M. Buda, A. Maki, and M. A. Mazurowski, “A systematic study of the class imbalance problem in convolutional neural networks,” *Neural Netw.*, vol. 106, pp. 249–259, Oct. 2018.

Celso Pereira received the M.Sc. degree in electronic and telecommunications engineering from the University of Aveiro, Aveiro, Portugal, in 2021. Since 2022, he has been working toward the Ph.D. degree in electrical and computer engineering with the Faculty of Engineering, University of Porto, Porto, Portugal. His research focuses on the intersections of autonomous driving, computer vision, machine learning, and robotics. In 2023, he actively contributed to the THEIA-Automated Perception Driving project, a collaborative initiative between the University of Porto and Bosch Car Multimedia Portugal.

Ricardo P. M. Cruz received the B.S. degree in computer science and the M.S. degree in applied mathematics from the University of Porto, Porto, Portugal, and the Ph.D. degree in computer science in 2021 with a special emphasis on computer vision and deep learning. Since 2015, he has been a Researcher with INESC TEC, Porto, working on machine learning with particular emphasis on computer vision. He is currently a Postdoctoral Researcher of autonomous driving under the ATLAS research project, a partnership between the University of Porto and Bosch Car Multimedia.

João N. D. Fernandes received the Ph.D. degree in civil engineering from the School of Engineering, University of Minho, Braga, Portugal, in 2021. He is currently a former Researcher with the Department of Electrical and Computer Engineering, Faculty of Engineering, University of Porto, Porto, Portugal. He is also a Researcher of artificial intelligence with INESC-TEC, Porto. Additionally, throughout his career, he has contributed to three research projects and supervised six master’s thesis and two internships. His research interests include optimization and deep learning.

João Ribeiro Pinto received the M.S. degree in bioengineering in 2017, and the Ph.D. degree in electrical and computers engineering from the University of Porto, Porto, Portugal, in 2022. From 2022 to 2023, he was a Senior Deep Learning Researcher with Bosch. He has coauthored more than 40 publications, supervised 16 M.S. students, contributed to six projects. His research interests include biometrics and well-being monitoring, especially within the context of intelligent vehicles. He was the recipient of the 2022 Max Snijder Award by the European Association for Biometrics.

Jaime S. Cardoso (Senior Member, IEEE) received the B.S. degree in electrical and computer engineering, the M.Sc. degree in mathematical engineering, and the Ph.D. degree in computer vision from the University of Porto, Porto, Portugal, in 1999, 2005, and 2006, respectively. He is currently a Full Professor with the Faculty of Engineering, University of Porto and a Researcher with INESC TEC, Porto. His main research interests include computer vision, machine learning, and decision support systems. From 2012 to 2015, he was the President of the Portuguese Association for Pattern Recognition, affiliated with the IAPR. Since 2011, he has been a Senior Member of IEEE. He serves regularly as a Project Evaluator for the European Commission.