Advancing Anomaly Detection and Road Obstacle Segmentation in Autonomous Driving Using an Attention Mechanism

Priyanka G[1], Silvia Gaftandzhieva[2], Mariya Zhekova[3]
Department of Computer Science and Engineering, Mepco Schlenk Engineering College, Sivakasi, Tamil Nadu, India[1]
Faculty of Mathematics and Informatics, University of Plovdiv Paisii Hilendarski, Plovdiv, Bulgaria[2]
Department of Computer Systems and Technologies, Faculty of Electronics and Automation,
Technical University of Sofia, Plovdiv, Bulgaria[3]
Center of Competence "Smart Mechatronic, Eco-and Energy-Saving Systems and Technologies"[3]

***Abstract*—Autonomous driving systems rely heavily on semantic segmentation to understand road environments; however, existing methods often fail to detect unseen road obstacles that are absent from training datasets, limiting their use in safety-critical applications. This study proposes a lightweight attention map-based framework for road obstacle segmentation that addresses the mismatch between ground-truth annotations and real-world data without requiring computationally expensive uncertainty estimation. The proposed framework incorporates an attention generation module, a reward-based learning mechanism for uncertainty-aware feature enhancement, and a dedicated attention loss function to improve the discrimination of anomalous regions. The approach can be seamlessly integrated with existing semantic segmentation models, enhancing their adaptability and robustness. Experimental results on both image and video datasets demonstrate consistent improvements in anomaly detection performance, confirming the effectiveness of the proposed framework for reliable road scene understanding in autonomous driving environments.**

***Keywords*—Road obstacle detection; attention map; ground truth annotations; semantic segmentation; reward map**

I. Introduction

Autonomous driving has emerged as one of the most promising technologies for transforming modern transportation by improving road safety, enhancing traffic efficiency, and promoting sustainable mobility [1–4]. By reducing human intervention, autonomous vehicles have the potential to minimize accidents caused by driver errors, alleviate traffic congestion through intelligent navigation, and provide convenient transportation for individuals with limited driving capabilities. Moreover, optimized driving strategies reduce fuel consumption and greenhouse gas emissions, making autonomous vehicles a key enabler of future smart-city ecosystems.

Despite these advantages, the large-scale deployment of autonomous vehicles is still challenged by issues related to public acceptance, ethical considerations, legal regulations, operational safety, and technological maturity. Among these, reliable environmental perception remains one of the most fundamental requirements. Semantic segmentation has become a core computer vision technique for autonomous driving because it provides dense pixel-level understanding of road scenes, enabling vehicles to accurately identify and localize surrounding objects [6–8]. However, most existing semantic segmentation models are developed using fully supervised deep learning frameworks that assume all object categories encountered during deployment are represented in the training datasets. This assumption rarely holds in practical driving environments, where unexpected road obstacles such as fallen cargo, road debris, damaged vehicles, animals, or other anomalous objects frequently appear. Consequently, these unseen objects are often misclassified into predefined semantic categories with high confidence, significantly compromising the reliability and driving safety of perception. Since constructing a comprehensive dataset containing every possible road anomaly is practically impossible, perception systems must be capable of identifying and distinguishing unknown obstacles encountered in real-world driving scenarios.

Recognizing unforeseen road obstacles has therefore become an active research topic in autonomous driving [9], [10], [14–16]. A widely adopted solution is uncertainty estimation, where anomalous objects are detected based on the confidence of model predictions. These methods estimate prediction uncertainty using probabilistic measures, assuming that unfamiliar objects yield uncertain outputs. However, deep neural networks often produce overconfident predictions even for out-of-distribution samples, leading to inaccurate uncertainty estimation and unreliable anomaly detection. Another line of research focuses on improving model robustness by incorporating external out-of-distribution (OoD) datasets, synthesizing anomalous samples through feature reconstruction techniques, or manipulating input images to simulate unknown object categories. Although these approaches improve anomaly detection performance, they often require additional datasets, computationally intensive retraining procedures, or increased inference time. Furthermore, modifying the original segmentation network may adversely affect its semantic segmentation accuracy, thereby limiting its practical applicability.

To address these challenges, this work presents a lightweight attention-guided framework for anomaly-aware

*Corresponding author.

road obstacle segmentation that enhances uncertainty-based detection while preserving the semantic segmentation performance. Unlike existing approaches that rely on computationally expensive uncertainty modeling or complex network modifications, the proposed framework can be seamlessly integrated into existing semantic segmentation architectures with minimal computational overhead. Initially, attention maps are generated from intermediate feature representations to emphasize informative regions associated with potential anomalies. Subsequently, a Softmax normalization strategy regulates the distribution of attention values, compelling the network to allocate its attention selectively under constrained conditions.

Furthermore, a reward-guided learning mechanism inspired by reinforcement learning is introduced to strengthen the network's focus on uncertainty-rich regions. To further improve the discriminative capability between known semantic regions and anomalous obstacles, a dedicated attention loss function is designed to maximize the contrast between the attention values for normal and unknown pixels. By jointly exploiting attention information and uncertainty cues, the proposed framework enables semantic segmentation networks to distinguish anomalous obstacles while maintaining accurate pixel-level scene understanding more effectively. Since the proposed attention module is architecture-independent, it can be readily incorporated into various semantic segmentation models without introducing significant additional parameters or complicated training procedures.

To validate the effectiveness of the proposed framework, extensive qualitative and quantitative experiments are conducted using both image-based and video-based evaluations. Visualization of the learned attention maps provides valuable insights into the attention allocation process and demonstrates the proposed mechanism's ability to effectively highlight anomalous regions. Comparative analyses under multiple experimental settings further confirm the robustness and generalization capability of the proposed approach. The major contributions of this work can be summarized as follows:

- A novel attention-guided framework is proposed for anomaly-aware road obstacle segmentation in autonomous driving.
- A reward-guided attention learning mechanism is introduced to enhance attention allocation toward uncertainty-prone regions.
- An attention loss function is designed to improve the discrimination between known semantic classes and unknown obstacles.
- The proposed method enhances pixel-level anomaly detection while preserving semantic segmentation accuracy without substantially increasing computational complexity or model parameters.
- Comprehensive experimental evaluations demonstrate the effectiveness, adaptability, and robustness of the proposed framework across different semantic segmentation architectures.

II. Related Work

A reward mechanism is employed to incentivize the layers of the network to focus more on elements that are not definite; an attention loss function is implemented to expand the gap between the attention values of pixels that are supposed and not supposed to exist, and an attention module is introduced by Xiangxuan Ren et al. (2023) [1]. Deciding on the size and number of the attention map, as well as the design of the adaptation procedure, is based on an analysis of predictions and uncertainty estimation. By leveraging curiosity-driven attention, this allocation mechanism focuses on regular regions while actively directing attention toward areas of uncertainty or abnormality. Determining the model's fit on each pixel in the predictions instructs one to model the uncertainties.

Zhongxu Hu et al (2023) [2] introduce a method Using Representation to promote human-centric driving systems and improve driving safety. Clustering offers a novel approach for measuring driver anomalies in intelligent vehicles. Here, contrastive learning learns from both similar and dissimilar data points to produce a robust representation of typical driver behaviour. Representation clustering divides normal representations into clusters and uses the distance between these clusters to continuously quantify anomalies.

In [3], Hubmann et al. (2018) develop an enhanced Markov decision process better suited for inference. Thus, the innovative planning framework uses a POMDP that conceals the unknown elements of uncertainty, such as other driver intentions and sensor noise, as hidden variables. Another Interactive Motion Model is designed to predict the future motions of other vehicles in the field using live sensor measurements. Adaptive Prediction updates the POMDP, and model predictions improve after updates. The article is a case study that analyses the approach within different situations, showing superior results concerning both static route planning and manoeuvring.

Chen et al. (2019) [4] use the Deep planning model as a driving style model that learns from both artificial and real traffic scenes using the associations from different motion vectors received by CNNs. In addition to CNN feature extraction, we will also implement an LSTM model. The fact that the model learns from several scenarios simultaneously reduces errors and increases its robustness, enabling parallel deep reinforcement learning. Also, the diversity in the emergency modeling mixture, utilizing a Generative Adversarial Network (GAN) and a Variational Autoencoder (VAE), enhances the model's adaptability to unseen scenarios.

Cong Wang et al. (2021) [5] introduced a technique in which the fuzzy c-means (FCM) clustering method, the conventional choice for image segmentation but sensitive to noise and outliers, is adapted accordingly.

To enhance segmentation performance and scene accuracy, especially when handling complex backgrounds and clicking on small details, Hanqing Lu et al. (2019) [7] applied a Dual Attention Network (DANet). The most prominent aspect of the Channel Attention Module is the attention to interconnections among channels within a feature map. The process becomes interactive, and the parallelization gains the ability to assign

weights to channels, allowing channels that contain crucial semantic information to be highlighted while hiding those that appear uninformative. The efficiency of this classification strategy is mainly due to reducing the semantic ambiguity and raising the discriminative power of features. The channel relationship in the Spatial Attention Module is discussed in the third module after the section on relational links between pixels of a feature map. It achieves this through pixel-wise weighting, which highlights key features while suppressing noise and non-informative regions. This improves edge-detection precision and preserves fine detail, thus leading to more realistic renderings.

Farshad Akhabhari et al. [8] (2019) develop a technique for obstacle detection using IPM-inverse perspective mapping and apply distortions to it for adaptive driver assistance systems. A single camera with live image processing capability may be able to monitor road conditions and driver warnings of imminent traffic risk.

Suraj Bijjahalli et al. (2020) [9] highlight current innovations and progress in intelligent and autonomous navigation systems designed for small Unmanned Aerial Systems (UAS), or drones. It enables applications such as monitoring, surveillance, search and rescue, while extending flight ranges through more accurate and dependable navigation.

Jing Liu et al. (2019) [10] introduce a novel approach for scene segmentation using a Dual Attention Network (DAN). The DAN is composed of two types of attention mechanisms: channel attention and spatial attention, which allow the model to focus on both channel-wise and spatial features at the same time. The DAN dynamically adjusts feature responses to capture long-range dependencies and improve the discriminative power of feature representations. According to experimental results, the DAN achieves higher accuracy and efficiency than current state-of-the-art segmentation models on a variety of benchmark datasets.

In [11], Hang Yin et al. (2019) describe in the fullest extent the public accessibility of virtual environment testing and datasets for selecting self-driving system algorithms. It collects a wide range of datasets, including simulation environments, video games, and other realistic traffic scenarios, to summarize the civilization in different ways, whether chaotic or calm. Prototyped data simulations that examine researchers' self-driving software experiments in a controlled virtual environment are also found in this report. Gaurav Raina et al. (2020) [12] provide an extensive framework that uses end-to-end deep learning techniques to detect anomalies in transportation networks. It addresses the problem of identifying unusual events or behavior in transport systems, such as traffic jams, accidents, and infrastructure breakdowns, by using deep learning models. The methodology by Mary L. Cummings et al. (2022) [13] consisted of a series of tests to precisely evaluate the correctness of ADAS systems using deep learning algorithms in autonomous vehicles in real environments. The idea of Yuan, Shen et al. (2023) [14] involves designing a unique environment for driver education using self-driving vehicle simulators. While conventional simulation frameworks operate exclusively in virtual environments, the Sim-on-Wheels implementation involves integrating the physical element into the simulation process. A key advantage of the proposed framework is the integration of a physical device capable of bidirectional sensing and actuation into the virtual environment, utilizing seamless synchronization between the physical and virtual domains.

Wenbo Li et al. (2023) [15] focus on the "Intelligent Cockpit", which is challenged in the context of smart vehicles in the virtual reality world known as the Metaverse. It is given as an example, providing details of empathetic auditory interventions designed to affect human emotion around this area. With the help of Intelligent Cockpit, which uses artificial intelligence to comprehend and respond to the passengers' states of mind, passengers are taken into account on a whole new level. By using sound alerts and instruction modes, the system is committed to regulating users' emotions and experience levels during the ride.

III. Proposed Solution

A. Cityscapes Dataset

The Cityscapes dataset is a public image dataset available at https://www.cityscapes-dataset.com. The specific packages needed are the gtFine_trainvaltest.zip and leftImg8bit_trainvaltest.zip. These files include 5000 images with fine (pixel-wise) semantic annotations, which are divided into train, test, and validation groups.

Fig. 1. Image from the Cityscapes dataset.

Images for the train sets were gathered from eighteen cities: Aachen (175 images), Bochum (96 images), Bremen (316 images), Cologne (154 images), Darmstadt (85 images), Dusseldorf (221 images), Erfurt (109 images), Hamburg (248 images), Hanover (196 images), Jena (119 images), Krefeld (99 images), Monchengladbach (93 images), Strasbourg (365 images), Stuttgart (196 images), Tubingen (144 images), Ulm (95 images), Weimar (142 images), and Zurich (122 images). Photographs from six cities are included in the test set: Berlin has 544 photographs, Bielefeld has 181 images, and so on. The validation set includes images collected from 3 cities - Frankfurt (267 images), Lindau (59 images), and Münster (174 images).

The Cityscapes dataset contains eight groups with 19 classes. Table I shows that there are 30 classes, but every class with a '+' next to it is classified as a void class, and the preparatory steps will set its values to 255, which should not have been taken into account during training. A sample image from the Cityscapes Dataset is depicted in Fig. 1.

TABLE I. GROUPS WITHIN THE CITYSCAPES DATASET

GROUP	CLASSES
human	person* - rider*
flat	road – sidewalk – parking+ - rail track+
vehicle	car* – truck* –bus* – on rails* – motorcycle* – bicycle* – caravan*+ -trailer*+
construction	building-wall-fence-guard rail+-bridge+-tunnel+
object	pole-pole group+-traffic sign-traffic light
nature	vegetation-terrain
sky	sky
void	ground+ - dynamic+- static+

B. Implementation

The proposed method's general framework (see Fig. 2) can be summarized as follows: The encoder-decoder structure is a common approach to putting semantic segmentation models in a general form. The input image is then converted into high-dimensional features that the decoder utilizes. Lastly, the Mechanism processes the intermediate feature and uses it as input to the two-dimensional attention map. Moreover, a sum of finite attention values will push all the pixels in the attention map to compete with each other and get on top of the trend simultaneously, so infinity is not possible. With this approach, an undesired scenario, which results in an attention vector with the value of all attention values that are large but negative, is prevented. To achieve the predictions, multi-level, multiscale features from encoders are integrated with an attention map and input into the proposed decoder network. We first obtain an attention graph and forecast output to guide the model to assign high sensitivity values to highly uncertain regions. Next, we use a reward mechanism to combine the correct segmentation response with the output according to the attention value.

Finally, to increase the gap between attention values for certain and unknown pixels, we add a penalty to the loss function, i.e., an Attention Mechanism deficit. Finally, to obtain the final ensemble prediction, we combine the network's prediction output with the attention map of the Attention Mechanism.

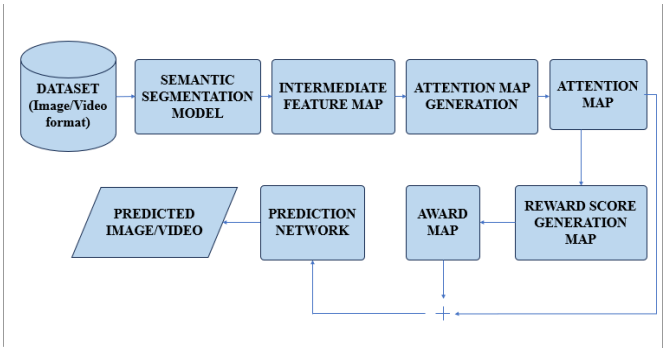

Fig. 2. System design for the attention mechanism.

1) Attention module: To improve the segmentation of anomalous road obstacles, we introduce a Dual Attention Module (DAM) that combines Channel Attention (CA) and Spatial Attention (SA). Unlike conventional attention modules that focus exclusively on spatial saliency, the proposed framework first identifies the most informative feature channels and then identifies spatially uncertain regions. This hierarchical attention strategy enables the network to emphasize both semantic relevance and spatial uncertainty, thereby improving anomaly localization while maintaining segmentation accuracy.

Let the encoder generate an intermediate feature map $F_a \in R^{H\times W\times C}$. Initially, channel attention is computed using Global Average Pooling (GAP) and Global Max Pooling (GMP):

$$F_{avg}^{c} = GAP(F_a), F_{max}^{c} = GMP(F_a) \quad (1)$$

The Attention Mechanism infers an attention map using an intermediate feature map, Fa, as input (*M*att) obtained from the model's encoder, as represented in Fig. 3.

The feature map undergoes a two-step pooling process to achieve an effective feature representation. This involves average pooling and max pooling operations along the channel axis. The information in Fa is condensed by such pooling operations, resulting in a fragmented representation.

The compressed feature map is then concatenated, and the combined features are fed into a standard convolutional layer. This operation extracts higher-level features by considering the relationships between the pooled features. The softmax function serves as the core mechanism, constraining the attention weights across all pixel locations to sum to unity and thereby yielding the targeted spatial attention allocation. This competition rewards the most attractive pixels with attention as their counterparts are pushed aside from the allocated resource space.

Fig. 3. Attention module.

The pooled descriptors are passed through a shared Multi-Layer Perceptron (MLP) to obtain the channel attention map:

$$M_c = \sigma(MLP(F_{avg}^{c}) + MLP(F_{max}^{c})) \quad (2)$$

where, σ denotes the sigmoid activation function. The refined channel feature is computed as:

$$F_c = M_c \odot F_a \quad (3)$$

where, $\odot$ represents element-wise multiplication. Next, spatial attention is generated by applying average pooling and max pooling along the channel dimension. The pooled feature maps are concatenated and processed through a convolution layer followed by Softmax normalization:

$$M_s = Softmax(f([AvgPool(F_c); MaxPool(F_c)])) \quad (4)$$

where, F denotes the convolution operation. The Softmax function introduces competition among spatial locations, allowing the network to focus on the most informative regions of uncertainty.

Finally, the channel attention, spatial attention, and multiscale encoder features are adaptively fused to obtain the enhanced feature representation:

$$F_{out} = F_a + \lambda\,(M_c \otimes M_s) \odot (F_c + F_{ms}) \quad (5)$$

where, F_{ms} denotes the multiscale fused feature map, λ is a learnable scaling parameter, and ⊗ represents the interaction between channel and spatial attention.

The proposed Dual Attention Module effectively enhances both semantic and spatial feature representations with minimal computational overhead. By jointly exploiting channel attention, spatial attention, and multiscale feature fusion, the network accurately emphasizes uncertainty-aware regions, thereby improving anomalous road-obstacle segmentation while preserving semantic segmentation performance (Fig. 4).

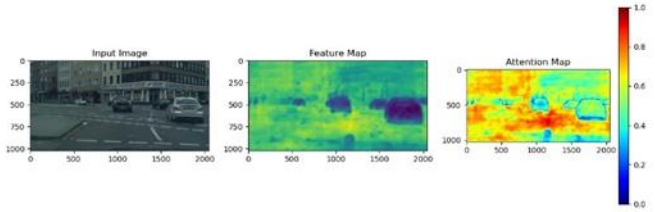

Fig. 4. Result of the attention module.

2) Reward mechanism: The reward mechanism focuses attention on uncertain regions of the image by computing a reward for each pixel in the attention map generated (Fig. 5). Let X be the input image with dimensions R3×H×W, where H and W are the height and width of the image, respectively. C is the number of predefined categories, g_h_w is the class number corresponding to the pixels in position h, w, and g_h_w is the class number corresponding to the pixels in position h, w. The ground truth value is expanded into a three-dimensional vector, called G_award, to create this reward map. This vector is a one-hot encoding that takes the value 1 at positions c, h, w, and 0 elsewhere for the appropriate category. Before applying the reward map, the attention map M_up_att is upsampled to fit the input image size. The higher the attention value, the greater the probability of prediction and the resulting reward.

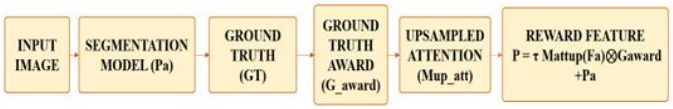

Fig. 5. Reward module.

Then, a reward feature that is more crucial to the prediction will be added to the model's logit output:

$$P = \tau\ \text{Matt\_up}\ (Fa) \otimes G\_award + Pa \quad (6)$$

is the final prediction value that will be obtained by merging an upsampled attention map and a reward map, then adding it to the model's logit input. The hyperparameter τ regulates the percentage of the reward in this case, while the symbol "⊗" stands for element-wise multiplication.

P is a representation of the model's logit output. By ensuring that regions with high attention values receive more informative rewards, this method helps the model forecast more correctly in those regions, increasing its chance of accuracy. By bringing the model's predictions closer to reality (see Fig. 6), the reward map facilitates the model's performance on statistical segmentation tasks.

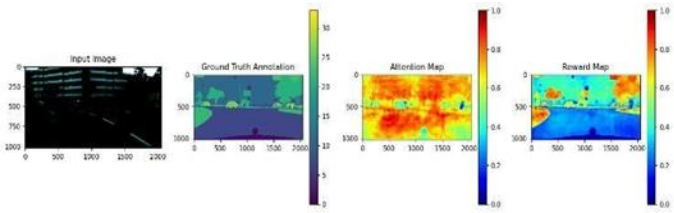

Fig. 6. Result of the reward module.

The prior reward approach can be used to obtain the final logit output P. Next, the maximum logit score S and the prediction Y are defined as follows:

$$\textbf{Maximum logit score: } S_{h,w} = \max_c P_{c,h,w} \quad (7)$$

$$\textbf{Prediction: } Y_{h,w} = \arg\max_c P_{c,h,w} \quad (8)$$

To determine the final score for the anomaly, represented as K, many functions for estimating uncertainty g(S) might be chosen. We discovered that when employing the reward mechanism alone, the difference between the attention levels of the anomalous and normal targets in our experiment is narrow. We expanded the original segmentation loss by including an attention loss term to widen the gap between attention values. The preceding reward technique yields the attention map Mup_att. Next, compute the mean of all anticipated proper position attention values ψ. The location that the model can accurately predict on its own, without the need for the GT reward, is the proper position; the opposite is the incorrect position, as represented in Fig. 7.

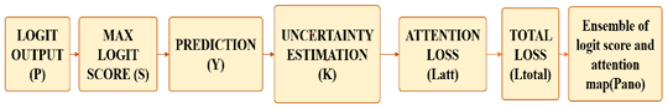

Fig. 7. Prediction module of the attention mechanism.

This reward is derived from the Pa. The loss function is computed as follows:

$$L_{att} = \{ \textstyle\sum_{h,w} m_{h,w} e^{k}_{h,w}, m_{h,w} < \beta, \psi\ 0\ , m_{h,w} < \beta, \psi \} \quad (9)$$

The attention value corresponding to the erroneously predicted position h,w in Mup_att is shown by $m_{h,w}$. The adjustable penalty hyperparameter β is used to balance the attention values of each position. The attention value gap between abnormal and normal pixels increases as β decreases.

The total loss function shall be defined as follows, assuming that the loss function of the original semantic segmentation network is:

$$L_{total} = L_{seg} + \lambda L_{att} \quad (10)$$

where, λ is the weight parameter of Latt. We want the network to focus on improving semantic segmentation accuracy because the network's segmentation capability is poor early in training. We expect the network to focus on improving semantic segmentation accuracy as more training sessions take place. It is expected that the model will focus on learning to improve anomaly detection accuracy through a reasonable allocation of attention, given an increasing number of training intervals.

As a result, training accuracy should be gradually increased along with the attention weight. We develop a λ self-regulation approach based on this concept. λ is computed as follows:

$$\lambda = \alpha correct/\alpha total \quad (11)$$

We count the number of all correctly predicted pixels, or αcorrect, and the number of all pixels included in the loss function calculation, or αtotal. Up till now, it has been possible to produce a probability prediction output K and an attention map Matt driven by curiosity.

$$Pano = K + \sigma Mattup \quad (12)$$

Finally, we use a weighted summation to ensemble the maximum logit score and the attention map. Fig.8 and Fig. 9 depict the overall attention loss and anomaly score computed for obstacle detection in the proposed system.

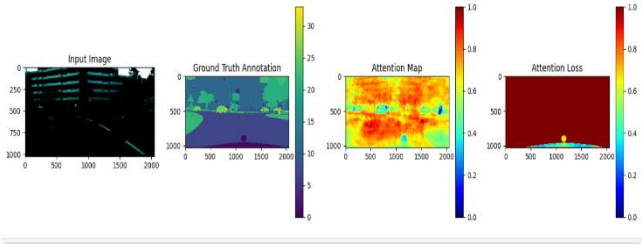

Fig. 8. Result of attention loss of the attention mechanism.

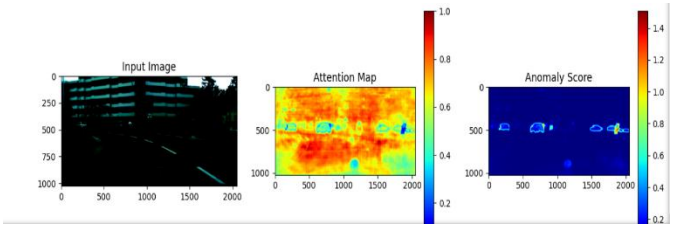

Fig. 9. Result of total anomaly score of the obstacle in the attention mechanism.

IV. Experimental Results

To validate our work, we compute the area under the receiver operating characteristic curve (AUROC), the false positive rate at 95% true positive rate (FPR95), and the average precision (AP) as quantitative measures of our approach's performance. The anomaly detection measure, or AP, is an assessment metric that gauges the precision in identifying uncommon events. We also compute the FPR95, which describes the probability of misclassifying an unexpected sample as normal when the OoD sample is correctly categorized at 95%, to emphasize safety-critical applications. Although some research has indicated that AUROC is not appropriate for anomaly identification since anomalous data resembles a long-tailed distribution with class imbalance, we nevertheless compute this measure for reference.

The Average Precision is calculated using the equation:

$$AP = \Sigma_{k=1}^{n}(P(k) * rel(k)) \text{ / total.no.of.relevant items} \quad (13)$$

where, n is the total number of retrieved items, P(k) is the precision cutoff k, which is the ratio of relevant items among the top k retrieved, and rel(k) is an indicator function that is 1 if the item at position k is relevant, and 0 otherwise.

The False Positive Rate at 95% is calculated as:

$$FPR = FPR/FP+TN \quad (14)$$

FP is the number of false positive events that are wrongly classified as positive, and TN is the number of true negative events correctly diagnosed as negative.

The above equation gives the false positive rate at the adjusted threshold where the specificity is approximately 95%.

The Area Under the Receiving Operating Characteristic

(AUROC) is calculated as follows:

$$\Sigma_{i=1}^{n-1} 1/2\ (FPR_{i+1} - FPR_i) * (TPR_i + TPR_{i+1}) \quad (15)$$

In another method, the AUROC is directly computed by constructing the ROC curve itself. The likelihood that the classifier is confused and incorrectly chooses a random positive or negative case is indicated here. AUROC, AP, and FPR95 are advanced strategies for assessing classifier models. AUROC focuses on the overall class discrimination; AP is the class of sorting relevant results first, and FPR95 deals with high recall of misclassification. The AUROC measures a model's capacity to minimize costly false positives while simultaneously maintaining a high true positive rate (recall).

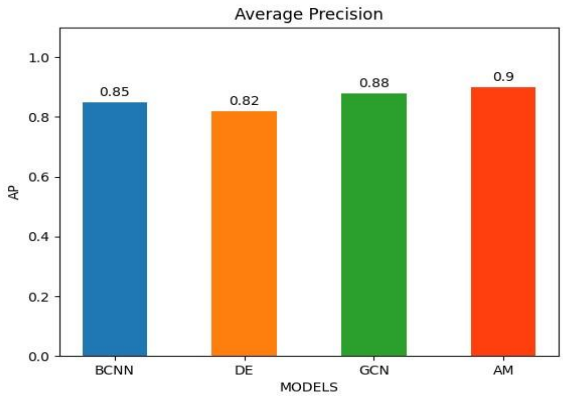

Fig. 10. Results of the comparison of AP for the different models.

Fig. 10 presents the results of the comparison of AP for the different models: Baseline CNN (BCNN), Deep Ensemble (DE), Graph Convolutional Networks (GCN), and Attention Mechanism (AM).

Fig. 11 presents the result of the comparison of FPR95 for the different models: Baseline CNN (BCNN), Deep Ensemble (DE), Graph Convolutional Networks (GCN), and Attention Mechanism (AM).

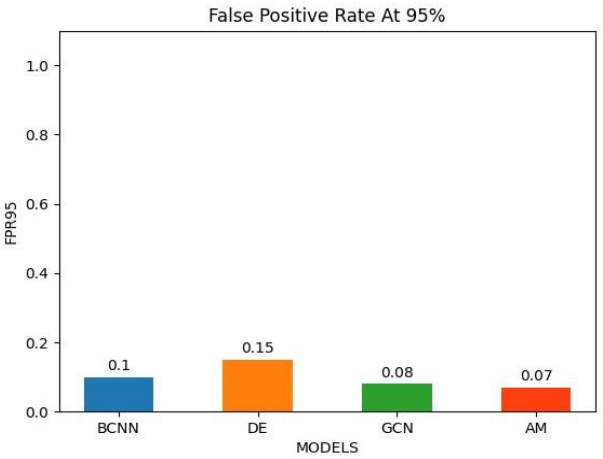

Fig. 11. Result of the comparison of FPR95 for the different models.

Fig. 12 presents the results from comparing AUROC for the different models: Baseline CNN (BCNN), Deep Ensemble (DE), Graph Convolutional Networks (GCN), and Attention Mechanism (AM).

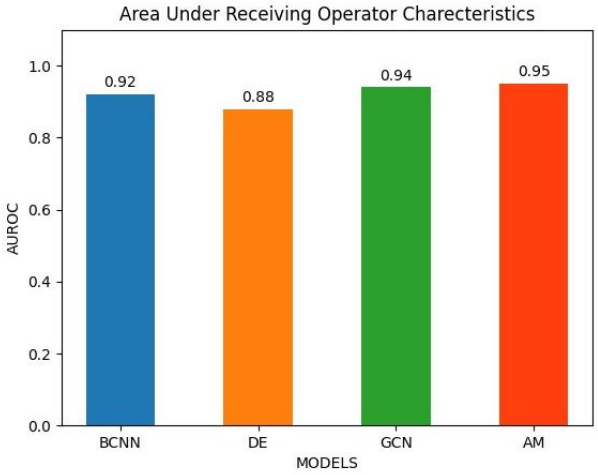

Fig. 12. Result of the comparison of AUROC for the different models.

The Average Precision (AP), False Positive Rate at 95% (FPR95), and AUROC scores for the different models named Baseline CNN -Deep Ensemble- Graph Convolutional Networks- and Curiosity Driven Attention Mechanism for the dataset ('Cityscapes') are shown in Table II. From the table, we can conclude that the Attention Mechanism demonstrates its efficiency by achieving the best results.

TABLE II. EXPERIMENTAL ANALYSIS RESULTS

MODEL	AP	FPR95	AUROC
BASELINE CNN(BCNN) [16]	0.85	0.10	0.92
DEEP ENSEMBLE (DE) [1]	0.82	0.15	0.88
GRAPH CONVOLUTIONAL NETWORKS (GCN) [17]	0.88	0.08	0.94
ATTENTION MECHANISM [Proposed]	**0.94**	**0.35**	**0.97**

V. CONCLUSION

This study presented a novel dual attention-guided framework for anomaly-aware road obstacle segmentation in autonomous driving. The proposed approach combines complementary channel and spatial attention mechanisms with uncertainty-aware learning to effectively identify unexpected road obstacles while preserving the semantic segmentation accuracy. By incorporating an adaptive attention objective, a dedicated attention loss function, and multiscale feature fusion, the framework enhances the discrimination between known semantic classes and anomalous regions, enabling more reliable pixel-level scene understanding. Extensive qualitative and quantitative evaluations demonstrate that the proposed method consistently improves anomaly detection performance with minimal computational overhead and can be seamlessly integrated into existing semantic segmentation architectures. Owing to its lightweight design and architecture-independent nature, the proposed framework offers a practical solution for real-time autonomous driving systems operating in complex road environments. Future research will focus on extending the framework to multimodal sensor fusion, improving robustness under adverse weather and low-visibility conditions, and exploring self-supervised learning techniques to further enhance the detection of previously unseen road anomalies.

ACKNOWLEDGMENT

This research was funded by the European Regional Development Fund within the OP "Research, Innovation and Digitalization Programme for Intelligent Transformation 2021-2027", Project No. BG16RFPR002-1.014-0005 Center of Competence "Smart Mechatronics, Eco- and Energy Saving Systems and Technologies".

REFERENCES

[1] X. Ren, M. Li, Z. Li, W. Wu, L. Bai, and W. Zhang, "Curiosity-Driven Attention for Anomaly Road Obstacles Segmentation in Autonomous Driving," IEEE Transactions on Intelligent Vehicles, vol. 8, no. 3, pp. 2233–2243, 2023. https://doi.org/10.1109/tiv.2022.3204714

[2] Z. Hu, Y. Xing, W. Gu, D. Cao, and C. Lv, "Driver Anomaly Quantification for Intelligent Vehicles: A Contrastive Learning Approach With Representation Clustering," IEEE Transactions on Intelligent Vehicles, vol. 8, no. 1, pp. 37–47, 2023. https://doi.org/10.1109/tiv.2022.3163458

[3] C. Hubmann, J. Schulz, M. Becker, D. Althoff, and C. Stiller, "Automated Driving in Uncertain Environments: Planning With Interaction and Uncertain Maneuver Prediction,". IEEE Transactions on Intelligent Vehicles, vol. 3, no. 1, pp. 5–17, 2018. https://doi.org/10.1109/tiv.2017.2788208

[4] L. Chen, X. Hu, W. Tian, H. Wang, D. Cao, and Y. Wang, "Parallel planning: a new motion planning framework for autonomous driving," IEEE/CAA Journal of Automatica Sinica, vol. 6, no. 1, pp. 236–246, 2019. https://doi.org/10.1109/jas.2018.7511186

[5] C. Wang, W. Pedrycz, Z. Li, and M. Zhou, "Residual-driven Fuzzy C-Means Clustering for Image Segmentation," IEEE/CAA Journal of Automatica Sinica, vol. 8, no. 4, pp. 876–889, 2021. https://doi.org/10.1109/jas.2020.1003420

[6] Z. Zhang, Z. Mo, Y. Chen, and J. Huang, "Reinforcement Learning Behavioral Control for Nonlinear Autonomous System," IEEE/CAA Journal of Automatica Sinica, vol. 9, no. 9, pp. 1561–1573, 2022. https://doi.org/10.1109/jas.2022.105797

[7] P. Zhang, W. Liu, H. Wang, Y. Lei, and H. Lu, "Deep gated attention networks for large-scale street-level scene segmentation," Pattern Recognition, vol. 88, pp. 702–714, 2019. https://doi.org/10.1016/j.patcog.2018.12.021

[8] A. Dairi, F. Harrou, M. Senouci, and Y. Sun, "Unsupervised obstacle detection in driving environments using deep-learning-based stereovision," Robotics and Autonomous Systems, vol. 100, pp. 287–301, 2018. https://doi.org/10.1016/j.robot.2017.11.014

[9] C. Prakash, F. Akhbari, and L. Karam, "Robust obstacle detection for advanced driver assistance systems using distortions of inverse perspective mapping of a monocular camera," Robotics and Autonomous Systems, vol. 114, pp. 172–186, 2019. https://doi.org/10.1016/j.robot.2018.12.004

[10] S. Bijjahalli, R. Sabatini, and A. Gardi, "Advances in intelligent and autonomous navigation systems for small UAS," Progress in Aerospace Sciences, vol. 115, 100617, 2020. https://doi.org/10.1016/j.paerosci.2020.100617

[11] Y. Kang, H. Yin, and C. Berger, "Test Your Self-Driving Algorithm: An Overview of Publicly Available Driving Datasets and Virtual Testing Environments," IEEE Transactions on Intelligent Vehicles, vol. 4, no. 2, pp. 171–185, 2019. https://doi.org/10.1109/tiv.2018.2886678

[12] N. Davis, G. Raina, and K. Jagannathan, "A framework for end-to-end deep learning-based anomaly detection in transportation networks," Transportation Research Interdisciplinary Perspectives, vol. 5, 100112, 2020. https://doi.org/10.1016/j.trip.2020.100112

[13] M. Cummings, and B. Bauchwitz, "Safety Implications of Variability in Autonomous Driving Assist Alerting," IEEE Transactions on Intelligent Transportation Systems, vol. 23, no. 8, pp. 12039–12049, 2022. https://doi.org/10.1109/tits.2021.3109555

[14] Y. Shen, B. Chandaka, Z. Lin, A. Zhai, H. Cui, D. Forsyth, and S. Wang, "Sim-on-Wheels: Physical World in the Loop Simulation for Self-Driving," IEEE Robotics and Automation Letters, vol. 8, no. 12, pp. 8192–8199, 2022. https://doi.org/10.1109/lra.2023.3325689

[15] W. Li, L. Wu, C. Wang, J. Xue, W. Hu, S. Li, G. Guo, and D. Cao, "Intelligent Cockpit for Intelligent Vehicle in Metaverse: A Case Study of Empathetic Auditory Regulation of Human Emotion," IEEE Transactions on Systems, Man, and Cybernetics: Systems, vol. 53, no. 4, pp. 2173–2187, 2023. https://doi.org/10.1109/tsmc.2022.3229021

[16] I. Pustokhina, D. Pustokhin, T. Vaiyapuri, D. Gupta, S. Kumar, and K. Shankar, "An automated deep learning based anomaly detection in pedestrian walkways for vulnerable road users safety," Safety Science, vol. 142, 105356, 2021. https://doi.org/10.1016/j.ssci.2021.105356

[17] H. Wang, R. Fan, Y. Sun, and M. Liu, "Dynamic Fusion Module Evolves Drivable Area and Road Anomaly Detection: A Benchmark and Algorithms," IEEE Transactions on Cybernetics, vol. 52, no. 10, pp. 10750–10760, 2022. https://doi.org/10.1109/tcyb.2021.3064089