IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 22 December 2023, accepted 11 January 2024, date of publication 15 January 2024, date of current version 29 January 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3354379

RESEARCH ARTICLE

Scenario-Based Segmentation: Traffic Image Segmentation by GNN Based Driver's Scenario

SEUNGWOO NHAM[1], (Graduate Student Member, IEEE), JINHO LEE[1], (Graduate Student Member, IEEE), SEONGRYUL YANG[2], JIHUN KIM[2], AND SHUNSUKE KAMIJO[3], (Senior Member, IEEE)

[1]Graduate School of Interdisciplinary Information Studies, The University of Tokyo, Meguro City, Tokyo 153-8505, Japan
[2]Beyless, Sujeong-gu, Seongnam-si, Gyeonggi-do 13449, Republic of Korea
[3]Interfaculty Initiative in Information Studies, The University of Tokyo, Meguro City, Tokyo 153-8505, Japan

Corresponding author: Seungwoo Nham (minamisyou@kmj.iis.u-tokyo.ac.jp)

This work was supported by JST SPRING under Grant JPMJSP2108.

ABSTRACT This paper introduces the Scenario-Based Segmentation Network (SBS-Net), which highlights significant advances in autonomous driving. Through the integration of the Scenario Enhanced Graph Neural Network (SE-GNN) and graph re-match modules into the existing semantic segmentation network model based on driver's cognition (DCSeg-Net), our approach optimizes the graph construction process. The SE-GNN module enhances the extraction of critical features and relations within diverse driving scenarios, elevating both accuracy and the system's adaptability to complex driving scenarios. The introduced graph re-match module refines classification discrepancies, significantly boosting segmentation accuracy and refining the understanding of autonomous driving scenes. Beyond technological enhancements, this work outlines the expansion and diversification of our original dataset, strengthening the learning capabilities of our model by including specific classification labels and incorporating a broader range of driving scenarios. The utility of the improved SBS-Net is demonstrated through superior performance in Graph Construction Accuracy and Intersection over Union (IoU) measures, as highlighted in our evaluation metrics. These advances underscore the practical applicability of scenario-based segmentation in real-world autonomous driving scenarios, enhancing overall scene comprehension capabilities. The developments presented signal substantial progress in the field of autonomous driving technology.

INDEX TERMS Driver's scenario, graph neural network, scene understanding, segmentation.

I. INTRODUCTION

Advancements in technology have fueled the evolution of autonomous driving, requiring more detailed and scenario-specific detections than earlier technologies to ensure safety. Traditional segmentation [1], [2], [3], [4] and detection [5], [6], [7], [8] methods have been effective to some extent; however, they often lack a comprehensive understanding of traffic scenarios. In response to this, we propose Scenario-Based Segmentation, a novel segmentation approach that takes advantage of the power of Graph Neural Networks [9], [10], [11], [12]. This system not only segments according to the category or object, but it understands complex traffic regulations, intersections, lanes, and potential risks. The objective of Scenario-Based Segmentation Network (SBS-Net) is to provide more accurate, holistic, and scenario-adaptive segmentation, thereby improving the safety and reliability of an autonomous car's operation. By incorporating scenario information into the segmentation process, SBS-Net transcends the limitations of traditional methods, offering an engineered approach to develop safe driving environments. The following sections detail the system structure, methodology, advantages over existing methods, and potential future applications.

To address the limitations of traditional methods, SBS-Net uses a scenario-based approach that considers specific traffic situations and regulations. The proposed system uses graph neural networks to capture the dependencies and relationships

The associate editor coordinating the review of this manuscript and approving it for publication was Po Yang.

between different elements in a traffic scenario. Scenario information is encoded in the form of a graph, where each object and its attributes are represented as nodes and their relationships are represented as edges. These connections allow SBS-Net to capture contextual information and make more accurate predictions. Furthermore, SBS-Net incorporates a robust object-oriented management module that can identify abnormal nodes and handle missed perception detections. This ensures that the segmentation results are reliable and consistent, even in challenging scenarios.

The proposed SBS-Net system presents numerous advantages over the existing methods. First, it establishes a comprehensive understanding of traffic scenarios by considering not only individual objects, but also their relationships and interactions. Second, SBS-Net incorporates scenario-specific information, enabling it to adapt and display superior performance in different traffic situations. Finally, SBS-Net amplifies the reliability of segmentation results by effectively managing perception failures and refining potential discrepancies in raw perception data. Collectively, these advances contribute to the overall safety and effectiveness of autonomous driving systems.

In future applications, SBS-Net can be used in various scenarios of autonomous driving such as urban environments, highways, and complex intersections. By accurately segmenting and understanding the traffic scene, SBS-Net can assist in tasks such as object detection, path planning, and decision-making for autonomous vehicles.

The paper is structured as follows:

In Section II, we explore related work in the fields of driver's scenario, segmentation, scene understanding, and graph neural networks, setting the foundation for discussions on our proposed model and methods.

Section III discusses the scenario-based segmentation network, our proposed solution. Here we elaborate on the mechanics and the components of SBS-Net, including the DC Seg-Net, the Scenario Enhanced graph neural network, and the graph re-match module.

In Section IV, we present the results of our experimental analysis. We discuss the data sets used in our experiment, followed by a detailed breakdown of the qualitative and quantitative evaluations of SBS-Net.

Section V provides a thoughtful discussion on the results obtained, the implications and potential avenues for future research.

Finally, in Section VI, we recapitulate our findings in our conclusion, highlighting the benefits and prospective impacts of our proposed model in the realm of autonomous driving and basic understanding of the scene.

II. RELATED WORK

The field of perception of autonomous driving has seen significant advances in recent years, and researchers have explored innovative approaches to improve the understanding of complex driving environments. In this section, we discuss the related work that has shaped the landscape of autonomous vehicle perception. The exploration is structured into three key areas: Segmentation Methods, Scene Understanding Methods, and Graph Neural Network Methods. These sections provide insight into the evolution of segmentation techniques, the application of graph neural networks for spatial modeling, and the development of comprehensive scene understanding frameworks for autonomous driving.

A. DRIVER's SCENARIO

While driving involves a variety of situations, from navigating around parked vehicles and adjusting to lane reductions, to detecting pedestrians near the vehicle and driving on narrow roads, each situation requires a different reaction from the driver, varying from immediate halting to continued caution. Existing research has predominantly focused on scenarios that require an immediate stop, leaving the scenarios where drivers need to maintain caution less explored.

This skew primarily stems from the common perception that scenarios requiring immediate halt, due to their immediate potential risk, are more crucial. However, the latter scenarios, although requiring a lesser degree of urgency, are equally important in embodying the true essence of autonomous driving: achieving navigation that mirrors human-like cautiousness and awareness.

Several existing studies, including TrafficNet [13], "Pattern Recognition for Driving Scenario Detection in Real Driving Data" [14], "Defining and Substantiating the Terms Scene, Situation, and Scenario for Automated Driving" [15], and GeoScenario [16] have made significant strides in understanding and representing different driving scenarios. Moreover, studies such as "End to End Learning for Self-Driving Cars" [17] and "Continuous control with deep reinforcement learning" [18] have explored advanced methods for driving decision making and control.

However, they have a common limitation that they do not fully address low-urgency scenarios in their research scope. For example, MPDM [19] and "Game-Theoretic Planning for Self-Driving Cars in Multivehicle Competitive Scenarios" [20] dive extensively into high-stakes decision-making problems but fall short in addressing scenarios with lower urgency. However, these scenarios that are perceived less urgently are vital for comprehensive autonomous driving and deserve comprehensive attention and systematic study.

Path planning and accident detection are essential elements in understanding a driver's scenario for autonomous driving systems. Path planning [21], [22] encompasses the optimal decision-making process required for efficient navigation, and accident detection [23], [24] is pivotal in ensuring safety by swiftly identifying potential obstructions or risks. Both of these components heavily rely on the effective interpretation and understanding of the driving scenario, necessitating advancements in our current frameworks. It's here that the value of scenario-based segmentation becomes pertinent. By improving the granularity and accuracy of scene comprehension, it could substantially enhance both path planning and accident detection in autonomous driving.

To accurately handle these less studied driver scenarios, a more detailed segmentation approach is necessary. Current studies, while successfully addressing situations that require abrupt responses, have not delved enough into developing segmentation methodologies equipped to discern the subtleties that mark these cautious scenarios. Therefore, there is an inherent need to direct future research towards these nuanced scenarios. Advancements in this realm would not only capture a more complete perception of autonomous driving, but would also contribute to the evolution of more holistic, context-aware autonomous driving systems.

B. SEGMENTATION METHOD

Semantic segmentation methods have become indispensable components of autonomous driving perception systems, playing a crucial role in accurately identifying and classifying objects within a given scene. Among these methods, deep neural networks, including Fully Convolutional Networks (FCNs) [1], U-Net [2], DeepLab [25], and You Only Look One-Level Pixelwise Prediction (YOLOP) [26], have emerged as leading approaches for semantic and instance segmentation tasks.

FCNs [1], introduced by Long et al., revolutionized the field by pioneering the concept of pixel-wise classification. This innovation allowed for the precise assignment of semantic labels to each individual pixel in an image, facilitating the capture of fine-grained details essential for discerning intricate elements within driving environments.

U-Net [2], proposed by Weng et al., further advanced semantic segmentation by introducing a U-shaped architecture. This unique design, which features a contracting path to capture context and an expansive path to high-resolution localization, has demonstrated exceptional performance in producing accurate and detailed segmentation results. The ability of U-Net to effectively incorporate contextual information while maintaining spatial accuracy makes it particularly well suited for the demands of autonomous driving scenarios.

DeepLab [25], developed by Chen et al., made strides in semantic segmentation by introducing dilated convolutions. This innovation enables the network to expand its receptive field without sacrificing resolution, capturing both local and global contextual information. DeepLab's ability to enhance the model's accuracy in delineating objects in diverse and complex scenes contributes significantly to achieving a comprehensive understanding of driving environments.

YOLOP [26], an extension of the You Only Look Once series, introduces a pixel-wise prediction approach for semantic segmentation. By combining the efficiency of YOLO with the precision of pixel levels, YOLOP represents a significant advance in real-time semantic segmentation tasks. Its ability to provide accurate predictions at every pixel efficiently makes it a promising candidate for various applications, such as autonomous driving, where real-time responsiveness is crucial.

DCSeg-Net [27], a novel method for semantic segmentation, has gained attention for its emphasis on multilevel feature extraction, tailored specifically for autonomous driving scenarios. DCSeg-Net, short for Driver's Cognition-Based Semantic Segmentation Network, is designed to address the intricate details within driving environments by integrating multi-level features to accurately identify drivable areas, intersections, and various elements that can impact safety.

The pixel-wise classification capabilities of these advanced deep neural networks, including DCSeg-Net, contribute significantly to providing a comprehensive understanding of intricate and dynamic environments in autonomous driving. The high-precision segmentation achieved through these methods plays a crucial role in supporting robust decision-making processes and ensuring the safety and efficiency of autonomous vehicles on the road. As the field continues to evolve, semantic segmentation remains a cornerstone, continually pushing the boundaries of what autonomous driving perception systems can achieve.

However, most existing segmentation methods focus solely on pixel-level analysis without considering contextual relationships between different objects in the scene. This limitation hinders their ability to infer higher-level semantics of scenes and to understand the overall driving scenario effectively.

C. SCENE UNDERSTANDING METHOD

Scene understanding plays a crucial role in autonomous driving systems as it involves recognizing different objects and elements in a particular environment and understanding their relationships spatially and semantically for proper navigation.

Several significant works have been proposed in this field. For example, Badrinarayanan et al. introduced SegNet [28], a deep learning model for understanding high-resolution semantic scenes. SegNet uses a combination of convolutional network architecture and pixel-wise classification to ensure effective and accurate scene understanding.

A different approach to scene understanding was proposed by Karim et al. They combined long- and short-term memory with fully connected networks to create LSTM-FCN [29] for semantic segmentation and scene parsing. This model has the ability to process temporal sequences, allowing it to understand dynamic scenarios in real driving situations. In another notable work, Chen et al. developed a methodology known as 3D Object Proposals for Accurate Object Class Detection (3DOP) [30]. This method extends scene understanding capabilities from 2D to 3D scenarios using stereo or monocular images to generate 3D proposals. This addition of depth and dimension improves the understanding of complex scenes.

IEEE Access

Liang et al. proposed an advanced method for scene understanding using a Graph Neural Network in their study, “Symbolic Graph Reasoning Meets Convolution” [31]. They incorporated symbolic reasoning with statistical machine learning, allowing autonomous driving systems to effectively comprehend more complex scenarios.

Although these methodologies have significantly advanced the field, a comprehensive framework that synergizes both segmentation techniques and scene understanding methods still needs development to better understand the driving environment.

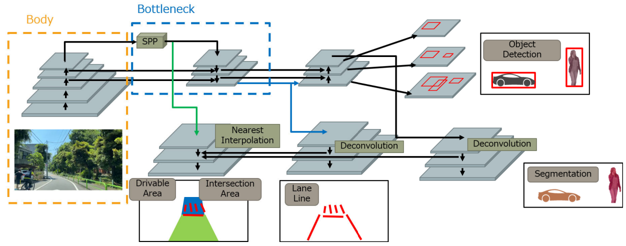

FIGURE 1. The DCSeg-Net [27] architecture is designed to identify intersection areas, in addition to the existing detection targets of YOLOP [26]. It is composed of a body (backbone), bottleneck, and three object detection and segmentation parts, which are based on the size of the target elements.

D. GRAPH NEURAL NETWORK METHOD

Graph Neural Networks (GNNs) are a widely used methodology in machine learning to capture and model complex relationships between entities. Scarselli et al. introduced GNNs [9] as an innovative approach to the handling of structured data, profoundly affecting perception in autonomous driving. The key idea behind GNNs is to understand and interpret graph-structured data, where entities are represented as nodes and their relationships as edges. By modeling these relationships, GNNs can extract crucial features and make meaningful interpretations.

In the realm of autonomous driving, the use of GNNs to discern relationships between various elements holds significant promise. Drawing inspiration from research on modeling IoT equipment with GNN, which focuses on learning relationships between sensor data and the simulation of equipment operation [32], the application of GNNs in the context of autonomous driving can offer valuable insights into understanding the intricate interconnections between diverse elements such as road conditions, vehicle dynamics and environmental factors. Using the principles and methodologies outlined in the study of IoT equipment modeling, the integration of GNNs into autonomous driving systems can potentially enhance the ability to capture and analyze complex relationships, contributing to the advancement of intelligent and adaptive autonomous driving technologies.

Graph Attention Networks (GATs) [33], introduced by Veličković et al., are an extension of GNNs that use a self-attention mechanism to weigh edges in a graph. GATs provide a more detailed analysis of the entities in the graph by assigning different levels of importance to different relationships. This fine-grained analysis allows the autonomous system to prioritize elements in a scenario and make more informed decisions.

Relational Graph Convolutional Networks (R-GNS) [34], proposed by Schlichtkrull et al., also contribute significantly in this domain. R-GNCS introduces relation-specific weight layers in the network that are capable of capturing multiple types of parallel edges or relationships in a scenario. This capability is vital in complex driving environments with a variety of interacting elements.

The focus of these methodologies on establishing a comprehensive understanding of the relational attributes within a scenario differentiates GNNs from other perception methods. The ability of GNNs to not only identify individual entities, but also understand their interactions helps improve the understanding of driving environments, providing autonomous systems with the necessary insights to make safer and more efficient decisions. However, the direct application of GNNs in a driving scenario requires an extension that integrates high-level semantics, which is the focus of the proposed scenario-based segmentation.

III. PROPOSED METHOD

The proposed method, termed scenario-based segmentation, is a new approach that integrates the strengths of both DCSeg-Net [27] and Graph Neural Networks to improve the understanding of driving scenarios.

DCSeg-Net, a Driver’s Cognition-Based Semantic Segmentation Network, forms the basis for the proposed methodology. It emphasizes multi-level feature extraction designed specifically for autonomous driving scenarios. DCSeg-Net accurately identifies drivable areas, intersections, and various elements that can impact safety by integrating layers of features.

To enhance the scenario understanding aspect of the segmentation, Graph Neural Networks are incorporated into this approach. GNNs are known for their ability to capture and model complex relationships between entities, providing a more detailed analysis of the entities by assigning different levels of importance to different relationships.

In the proposed method, the Graph Neural Network is applied to the segmentation network to facilitate more nuanced scene understanding. By modeling the relationships among different objects detected in the segmentation, GNNs map the context of the scene, taking into account the relationships between different entities and integrating this information to gain high-level semantics of the scene.

This combination of DCSeg-Net and graph neural networks allows the proposed scenario-based segmentation method not only to accurately identify individual entities such as drivable areas and intersections but also to understand their interactions. Understanding these interactions is crucial to ensuring the safety and efficiency of autonomous vehicles.

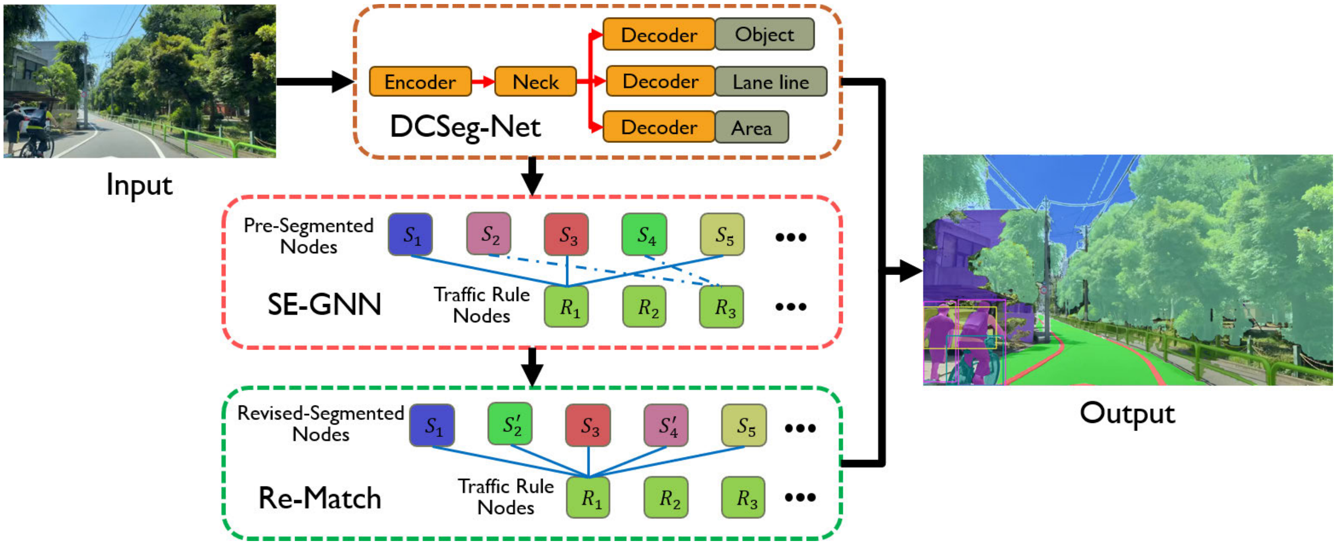

FIGURE 2. Comprehensive diagram of the Scenario-Based Segmentation Network. The network consists of three primary components. The first, DCSeg-Net [27], provides the initial semantic segmentation. The second, the Scenario Enhanced Graph Neural Network module, refines the segmentation from the DCSeg-Net by incorporating the context of driving scenarios to extract intricate features and relations. The final component, the Graph Re-Match module, computes the loss from the refined labels and minimizes it. This loss correction guides DCSeg-Net to learn the correct labels, enabling more accurate scene understanding and improved segmentation for autonomous driving applications.

A. DC SEG-NET: DRIVER's COGNITION-BASED SEMANTIC SEGMENTATION NETWORK

DCSeg-Net [27], or Driver's Cognition-Based Semantic Segmentation Network, is a cornerstone in the proposed scenario-based segmentation method. Marked by its proficiency in providing detailed analysis of road scenes, DCSeg-Net specializes in autonomous driving scenarios. The paper on this topic was presented at the IEEE International Conference on Intelligent Transportation Systems in 2023, held in Bilbao, Spain. The DCSeg-Net model plays an important role in understanding the cognitive process of drivers and effectively contributes to the development of segmentation methodologies in autonomous driving systems.

DCSegNet's strength primarily lies in its emphasis on multi-level feature extraction. This trait allows the network to identify crucial components and features within driving scenes, such as accurately recognizing drivable areas, detecting potential obstructions and intersections, and other details that are imperative for autonomous driving.

Multitiered feature extraction is executed through the integration of different layers of features in a multiscale approach. DCSeg-Net's advanced feature integration enables it to provide a robust and detailed segmentation that extends beyond focusing on individual entities within the scene to comprehensively understanding the full driving environment.

DCSeg-Net, a cutting-edge network architecture designed for autonomous driving perception, integrates key components from state-of-the-art models to achieve superior performance. Inspired by the YOLOP [26] model, DCSeg-Net uses a modified version of YOLOP to detect granular areas and objects, thereby enhancing its understanding of complex traffic scenes, such as opposite lanes and intersection areas. However, instead of using YOLOP's backbone YOLO (You Only Look Once) [8] as the backbone, DCSeg-Net utilizes CSPDarknet [35], a feature extraction model that represents superior object detection by dealing with the issue of gradient duplication. The bottleneck is made up of modules SPP (Spatial Pyramid Pooling) [36] and FPN (Feature Pyramid Network) [37], which generate and combine features on different scales and semantic information. This comprehensive approach allows DCSeg-Net to perform object detection/segmentation, drivable area segmentation, lane segmentation, opposite lane segmentation, and intersection segmentation tasks simultaneously, positioning it as a state-of-the-art solution for autonomous driving perception.

Therefore, DCSeg-Net is an essential part of the proposed scenario-based segmentation method. Its ability to discern critical features and elements within driving environments helps establish a nuanced understanding of the scene, which is essential to ensure the safety and efficiency of autonomous vehicles. By forming the basis of the proposed method, DCSeg-Net contributes to enabling a more improved understanding of driving scenarios, thus laying the groundwork for further scenario analysis through the integration of Graph Neural Networks.

In conclusion, the DC Seg-Net architecture, inspired by human cognition, forms a foundational element of our approach to scenario-based segmentation. It represents an integral step towards replicating a driver's perception in an autonomous driving system. This network facilitates a better understanding of the environment by distinguishing various elements within a driving scene, such as lane markings and potential obstacles. The integration of such cognition-based methods into autonomous vehicles is pivotal as it paves the

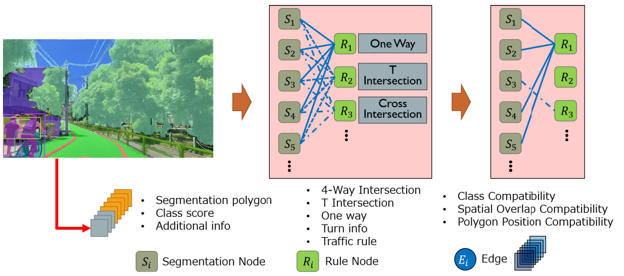

FIGURE 3. In-depth illustration of the Scenario Enhanced Graph Neural Network module. The SE-GNN operates by constructing a graph network, which includes two types of nodes: segmentation nodes derived from DCSeg-Net [27] and traffic rule nodes based on scenario context. Edges of the graph represent the relationships between these two types of nodes. This graph, thus, offers a structured representation of the driving scenario, laying the groundwork for subsequent stages of the model to refine segmentation and improve scene understanding in autonomous driving applications.

way for more accurate and reliable interpretation of driving scenarios, directly contributing to enhanced decision-making in complex traffic situations.

B. SCENARIO ENHANCED GRAPH NEURAL NETWORK MODULE

The proposed method for scenario-based segmentation, designed to complement DCSeg-Net and expand its capabilities, integrates a scenario-enhanced graph neural network module. This component improves segmentation information with specific scenarios, leading to a more accurate interpretation of real-world driving environments.

The Graph Neural Network consists of two types of nodes: segmentation information nodes that depict the output of the DCSeg-Net, containing details about segmented polygons and class scores; and rule nodes that encompass scenario-specific rules related to driving situations such as intersection types, traffic regulations, and input pertaining to the ego vehicle's turn signal state. The edges of the network represent the connections between these nodes and capture the interactions between segmented entities and traffic rules in a driving context. These edges provide contextual information that enables the system to comprehend the significance and connection of individual elements within the complete driving environment. By integrating both node types and their associations, the Graph Neural Network module can model and assess the intricate interplay among segmented scene elements and traffic regulations. This analysis goes beyond the recognition of individual components and understanding their semantic interactions. In summary, by leveraging DCSeg-Net's [27] capabilities, the Scenario Enhanced Graph Neural Network module combines segmented scene particulars with high-level scenario data and rule sets to achieve a more sophisticated comprehension of scenarios, thereby enabling more intelligent decisions for autonomous vehicles based on contextual awareness.

To represent the relationships between the segmentation nodes and the traffic rule nodes in our Graph Neural Network, we can assign the letter S to denote a segmentation node and the letter R to a traffic rule node. The edge weight W_{ij} between a segmentation node i and a traffic rule node j can be calculated using a sigmoid function applied to their relationship score:

$$W_{ij} = \frac{1}{1 + e^{-\beta \cdot C_{ij}}} \tag{1}$$

Here, e is the base of the natural logarithm, and β is a parameter that controls the sensitivity of the edge weights to the compatibility function. The function C_{ij} could be a measure of the precision of the segmented polygons with traffic regulations or the characteristics of the intersections.

The compatibility relationship score C_{ij} is determined using segmented scene attributes and high-level scenario data:

$$\begin{aligned} C_{ij} = w_1 &\cdot \mathrm{ClassCompatibility}(S_i, R_j) \\ &+ w_2 \cdot \mathrm{SpatialOverlapCompatibility}(S_i, R_j) \\ &+ w_3 \cdot \mathrm{PolygonPositionCompatibility}(S_i, R_j) \end{aligned} \tag{2}$$

Here:

- *ClassCompatibility*(S_i, R_j) measures the compatibility of the classes assigned to the segmentation node S_i and the traffic rule node R_j.

- *SpatialOverlapCompatibility*(S_i, R_j) measures the spatial overlap compatibility between the segmented polygon of S_i and the expected region specified by R_j.

- *PolygonPositionCompatibility*(S_i, R_j) measures the compatibility of the polygon position between the segmented polygon of S_i and the expected position of the polygon of R_j.

- w_1, w_2, w_3 are weight coefficients that you can adjust based on the importance of each factor.

The edge connection can be binary, indicating whether there is a connection between a segmentation node S_i and a traffic rule node R_j. Define this connection using a threshold θ in the relationship score:

$$E_{ij} = \begin{cases} 1 & \text{if } C_{ij} \geq \theta \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

The Scenario Enhanced Graph Neural Network (SE-GNN) module utilizes the provided formulas to effectively model and assess the intricate connections between segmented scene elements and traffic rules. This empowers the system not only to identify individual elements, but also to understand their interactions and relationships within a driving scenario, thereby facilitating more informed decisions for autonomous vehicles based on contextual awareness.

This approach enables SE-GNN to accurately prioritize relationships and adjust feature representations for each node by considering both local segmentation node features and their corresponding rule node relationships. As a result, it fosters contextual awareness and robust significance, leading to improved understanding of driving scenarios. This representation is crucial to enabling the precise interaction

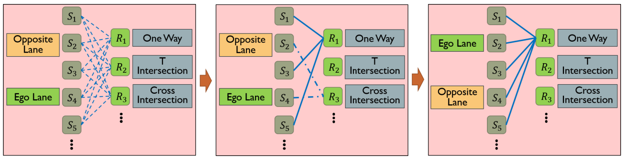

FIGURE 4. **Detailed schema of the Graph Re-Match module. This component operates by minimizing the losses associated with refined label predictions. The refined labels, generated by the previous stages, are compared with the target labels, and the differences are quantified as losses. These losses are then minimized using optimization techniques, and the corrected labels are used to guide the DCSeg-Net [27] for reclassification, thus refining the overall segmentation and achieving more accurate scene understanding in autonomous driving applications.**

of autonomous vehicles within complex traffic scenarios through the SE-GNN module.

In summary, the Scenario Enhanced Graph Neural Network Module is a key advancement in understanding complex driving scenes. It processes the relationships between different road elements and traffic participants, thereby enhancing the predictive capabilities of the system. By integrating SE-GNN, we ensure that the autonomous driving model not only recognizes objects but also comprehends the scene structure, enabling more intelligent and anticipatory driving decisions in real-time. This module is crucial for adapting to dynamic driving conditions and improving overall traffic safety.

C. GRAPH RE-MATCH MODULE

The "Graph Re-Match Module" is a crucial component that rectifies the misconnections between the segmentation nodes and the traffic rule nodes, optimizing the scenario-based segmentation for autonomous vehicles. This improvement module aims to reassess and re-match the relationships between the segmentation and traffic rule nodes.

During the initial scenario-based segmentation facilitated by the DCSeg-Net [27] and SE-GNN module, some segmentation nodes might not align accurately with their corresponding traffic rule nodes due to the complexity of real-world driving scenarios. These misaligned nodes can result in discordance with actual traffic rules, which could compromise the safety and precision of autonomous decision-making. The "Graph Re-Match Module" rectifies this issue by implementing an algorithm that reevaluates the 'fit' compatibility between the segmentation nodes and rule nodes based on traffic rules' logic. Each segmentation node, denoted as S_i, is initially assigned a class $Class_k$. The class $Class_k$ represents a category such as road, vehicle, pedestrian, etc., in which the segmentation node S_i is classified. For instance, if a certain segmented area in the scene has been identified as a pedestrian, the segmentation node representing that area is assigned the class 'pedestrian'.

The Graph Re-Match module plays an integral role in the overall function of the SBS-Net by providing feedback to the DCSeg-Net. The output from the Graph Re-Match module is used to refine the predictions from the DCSeg-Net in subsequent iterations. The refined labels provided by the Graph Re-Match module are compared to the target labels, and any loss is minimized during future training processes. This feedback mechanism allows DCSeg-Net [27] to progressively learn and improve its classification accuracy, thereby enhancing the overall system's performance in segmenting and understanding different driving scenarios.

The Graph Re-Match Module can change the class $Class_k$ of a segmentation node S_i, given an inconsistency between the two nodes. Let the updated class $Class'_k$ be the class that corresponds to the maximum relationship score with any traffic rule node R_j:

$$Class'_k = \mathrm{argmax}_{Class_k} C_{ij}(S_i, R_j) \tag{4}$$

This implies that the class $Class_k$ of the segmentation node S_i will be updated to $Class'_k$, resulting in the maximum relationship score C_{ij} with any traffic rule node R_j. This will further ensure that the segmentation node's class is in a better alignment with real-world traffic rules.

In addition to adjusting the class of the segmentation node, the module can also adjust the connection between the segmentation node and the traffic rule node. This can be carried out based on a threshold evaluation process θ similar to equation 3.

The revised graph, after the re-matching process, aligns more accurately with the actual traffic rules and represents a significantly improved understanding of the scene. Here, the edge E_{ij} symbolizes the existence of a connection between the segmentation node S_i and the traffic rule node R_j. By adjusting the class or connections of the segmentation node S_i, the module can create associations that match the traffic rules more accurately, improving overall precision in understanding the scene.

Consequently, the Graph Re-Match Module fosters that these improvements, in the class and connections of segmentation node S_i, contribute a high degree of adaptivity and precision in autonomous driving. These refinements implemented by the module ensure a more contextually accurate interaction of autonomous vehicles within complex traffic scenarios, underlying a promising future for autonomous driving.

To sum up, the Graph Re-Match Module is designed to refine the initial segmentation results from previous network layers. It acts like a quality check, identifying and correcting any segmentation errors. This module enhances accuracy by ensuring that the final output more closely aligns with the actual scenario, leading to more precise interpretations of the driving environment by the autonomous system. Its role is to act as a safeguard, reducing misclassifications that could affect the vehicle's decision-making process.

IV. EXPERIMENTAL RESULTS

The effectiveness and efficacy of the proposed graph re-match module were evaluated by a series of tests and comparisons. To provide a comprehensive assessment, our evaluations are categorized into two sections: a qualitative evaluation through visual outcomes and a quantitative

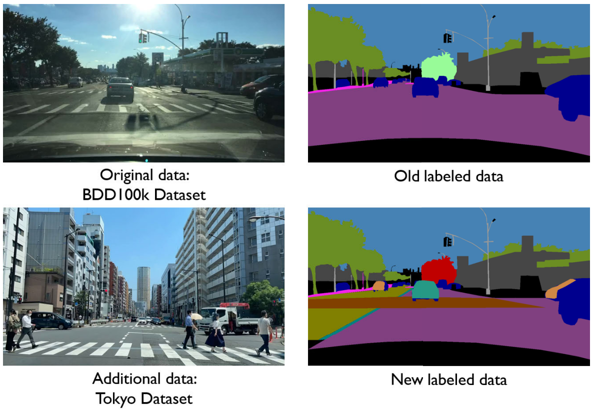

FIGURE 5. An example of four data is shown from top to bottom, left to right: the BDD100k [38] dataset, the Tokyo dataset which was collected by us, the old labeled data, and the new labeled data. The new labeled data has extra categories, such as the front and back of a vehicle, an intersection area, an opposite lane area, and graphs.

evaluation through distinct metrics. This section outlines the data set used for the train and testing, followed by the results obtained through both evaluations.

A. DATASET

Our research uses the BDD100K [38] dataset, which is well known for its wide range of information for autonomous driving studies. The dataset, however, did not include labels for more complex subjects, such as opposite lanes and intersection zones. To counteract this, we enriched the data set by adding these additional classifications.

More than 1000 images from the BDD100K dataset were sampled and labeled with these added classes, delineating vehicle front and rear, intersection areas, and areas of opposing lanes separately. This stratification enabled us to gather detailed information about the opposing lane during the learning process.

To augment our data range, we gathered test data from Tokyo, which also includes the newly included class labels. We specifically focused on incorporating images of vehicles turning at intersections because of the rarity of such data. Fig. 5 shows examples of our training data, test data, original labeled data, and newly categorized data, in order from top left to bottom right. These images visually represent the diversity and detailed segmentation in our dataset.

Furthermore, we develop distinct graph labels for our graph module to test the capabilities of the graph rematch module, intentionally including DCSeg-Net failure cases [27] to assess its refinement functionality. This extended and meticulously labeled data set serves as the basis for a more comprehensive scenario-based learning approach that improves the scene comprehension capabilities of autonomous driving systems.

B. QUALITATIVE EVALUATION

A powerful asset to further establish improvements in SBS-Net performance is qualitative evaluation. This evaluation, carried out through visual illustrations, vividly manifests the correction of segmentation errors against traditional methods.

Fig. 6 provides an illuminating overview of these enhancements. Visually compares the results of DCSeg-Net [27], a conventional model, with our proposed SBS-Net. The segmentation errors evident in the DCSeg-Net results are effectively resolved and refined in the images produced by the SBS-Net. Such a direct comparison accentuates the proficiency of SBS-Net and its superior handling of various complex circumstances. Provides a comprehensive scope into the overall performance of the model, reflecting its ability to not only correct specific errors, but also its overall efficacy in generating precise and accurate scene segmentation. These results offer a quantifiable and visual testimony to the enhanced capabilities of SBS-Net in scene understanding and corrections of segmentation errors for autonomous driving. This, coupled with the quantitative evaluation, offers a well-rounded evaluation of the impressive performance of the SBS-Net.

C. QUANTITATIVE EVALUATION

More objectively, we evaluated the performance of our proposed method using four metrics.

Drawing inspiration from the robust methods in the paper "Robust Construction of Spatial-Temporal Scene Graph Considering Perception Failures for Autonomous Driving" [39], the metrics utilized in our quantitative evaluation consider potential perception failures. These evaluation metrics offer a compelling means to assess the effectiveness and accuracy of our scenario-based segmentation model, thus ensuring the model's robustness in autonomous driving contexts.

1) COMPARISON OF INTERSECTION OVER UNION

For the first metric, we evaluate the Intersection over Union (IoU), focusing on lane lines and drivable areas across various intersection types, including "turn" and "go straight" scenarios. "Turn" scenarios refer to instances where the ego vehicle is near an intersection and is attempting a right or left turn. On the other hand, "go straight" scenarios correspond to instances where the vehicle proceeds directly ahead without making a turn. The IoU measures the overlap between the model's predicted segmentation and the ground truth, with a higher IoU signifying more accurate segmentation. By precisely focusing on the lane lines and drivable areas, which are critical for autonomous navigation, and comparing the IoU performance between DCSeg-Net [27] and our proposed SBS-Net, we can observe discernible improvements. Through the integration of these metrics, we can achieve a comprehensive evaluation of the model performance from multiple key angles. Table 1 shows this result.

The IoU is also compared to various traditional methods for both lane lines and drivable areas in Tables 2 and 3. This

FIGURE 6. Visual comparison of the segmentation results between DCSeg-Net [27] and SBS-Net. The image on the left shows the outcome of DCSeg-Net, where mis-segmentation is noticeable during a turn at a T-Intersection. The image on the right represents the refined segmentation result from SBS-Net, which accurately corrects the initial segmentation errors. This comparison demonstrates the effectiveness of SBS-Net in improving scene understanding in autonomous driving applications.

TABLE 1. The quantitative evaluation IoU for drivable area segmentation. Comparison for each scenario with DCSeg-Net [27].

Scenario	IoU of DCSeg-Net [27]	**IoU of SBS-Net**
One Way	95.2	**97.4**
T-Intersection (Turn)	67.7	**81.1**
Cross-Intersection (Turn)	56.8	**77.9**
Cross-Intersection (Go Straight)	91.2	**96.2**

TABLE 2. The quantitative evaluation IoU for drivable area segmentation. Comparison with traditional methods.

Network	IoU
MultiNet [40]	71.6
DLT-Net [41]	71.3
PSPNet [42]	89.6
YOLOP [26]	91.5
YOLOP v2 [43]	93.2
DCSeg-Net [27]	94.3
SBS-Net	**96.1**

TABLE 3. The quantitative evaluation IoU for lane line segmentation. Comparison with traditional methods.

Network	IoU
MultiNet [40]	71.6
DLT-Net [41]	71.3
PSPNet [42]	89.6
YOLOP [26]	91.5
YOLOP v2 [43]	93.2
DCSeg-Net [27]	94.3
SBS-Net	**95.7**

comparison further establishes the advantage of our proposed model, SBS-Net, over other conventional approaches like MultiNet, DLT-Net, PSPNet, and even YOLOP variants. SBS-Net consistently shows elevated IoU scores, highlighting its superior accuracy in scene segmentation, crucial for autonomous driving. This general comparison allows us to demonstrate the comprehensive and multifaceted performance of our model.

TABLE 4. Graph Construction Accuracy (GCA). Values of GCA across different types of intersections, demonstrating the efficacy of SBS-Net in accurately constructing traffic graphs.

Type of Intersection	GCA [%]
One Way	93.6
T Intersection	91.3
Cross Intersection	88.7

TABLE 5. Positive Refinement Ratio (PRR). PRR values demonstrating the effectiveness of SBS-Net in correcting segmentation errors across various intersection types.

Type of Intersection	PRR [%]
One Way	73.2
T Intersection	82.7
Cross Intersection	80.2

2) GRAPH CONSTRUCTION ACCURACY

Graph Construction Accuracy (GCA) is a key metric in our evaluation, as Table 4. It refers to the percentage of valid information in the constructed SE-GNN compared to the ground truth of the scene. Essentially, this metric assesses the way well the constructed SE-GNN aligns with the factual data of the scene. High GCA values indicate a higher correlation between the interpretation of the SE-GNN module and the accurate representation of the scene. Thus, it measures the efficacy of segmentation nodes and traffic rule nodes interpretation.

3) POSITIVE REFINEMENT RATIO

This metric is an indicator of the model's ability to effectively make accurate adjustments. The Positive Refinement Ratio (PRR) is the ratio of adequately refined nodes to the total perceived failure nodes in the scene. A higher PRR means that the module has successfully identified and amended misaligned node classifications, demonstrating its effectiveness in refinement. Table 5 displays this result.

4) NEGATIVE REFINEMENT RATIO

For the final, this metric refers to the ratio of inappropriately adjusted nodes to total refined nodes. In essence, the Negative

TABLE 6. Negative Refinement Ratio (NRR). NRR values depicting the precision of the adjustment process during the refinement phase for different types of intersections.

Type of Intersection	NRR [%]
One Way	1.5
T Intersection	5.3
Cross Intersection	4.6

Refinement Ratio (NRR) is a measure of misalignment introduced during the refinement process. A lower NRR signifies a higher degree of precision in the adjustment process during the graph re-match module, implying that it successfully circumvented unnecessary or inappropriate modifications. Table 6 displays this result.

V. DISCUSSION

In the course of our research, significant insights on scenario-based segmentation for autonomous driving were gained, especially in addressing the weaknesses of DCSeg-Net [27]. The newly implemented Graph Re-Match Module has proven successful in refining misclassification, as demonstrated by our evaluation metrics, notably the Positive Refinement Ratio and Negative Refinement Ratio.

The careful expansion and additional depth of our data set have also contributed remarkably to the efficacy of our system. By introducing detailed classification labels and incorporating diverse driving scenarios such as intersection turns, we managed to further refine and validate our model's learning capabilities.

Despite these improvements, some challenges were encountered. For example, handling rare or unique driving scenarios still requires further optimization. These are conditions that are not well represented in our current dataset, emphasizing the need for continuous expansion and diversification of the data set.

Considering the results of our quantitative evaluation, there is notable room for improvement in Graph Construction Accuracy and IoU metrics. Although our model overcomes its predecessor in accuracy, the objective of achieving superior performance in understanding autonomous driving scenes requires continuous enhancement in these areas.

In summary, scenario-based segmentation has shown promising potential in improving the scene comprehension capabilities of autonomous driving systems. Further research, data set refinement, and optimization of our system are still required to push our promising preliminary results toward concrete real-world applications.

VI. CONCLUSION

In this study, our objective was to provide significant information on the application of scenario-based segmentation for autonomous driving. Our focus centered on the improvements over the DCSeg-Net [27], involving the addition of the SE-GNN and the Graph Re-Match modules. These novel components have proven instrumental in rectifying misclassifications, taking us a substantial step forward towards precise scene understanding for autonomous driving.

The SE-GNN module has enhanced our model's ability to extract critical features and relations in the scenario, furthering its comprehension and decision-making capabilities. In contrast, the Graph Re-Match module, with its ability to refine misclassifications, has increased the accuracy of segmentation, yielding promising results in our evaluation metrics.

Our enhancement and diversification of the original dataset, by adding specific labels and capturing a range of driving scenarios, have significantly improved the learning capabilities of our model. Although these adjustments have improved model performance, addressing rare or unique driving scenarios remains an ongoing challenge, revealing the need for further data set extension and optimization.

Substantial room for improvement remains, as evidenced by our evaluation metrics, specifically in Graph Construction Accuracy and IoU measurements. Despite superior accuracy over its predecessor, our model requires continual refinements and adjustments to achieve excellent performance in its interpretation of autonomous driving scenes.

Ultimately, the potential of scenario-based segmentation to enhance the scene comprehension capabilities of autonomous driving systems is immense. However, to convert our preliminary results into practical real-world applications, continuous research efforts, data set enhancements, and further system optimizations are necessary.

ACKNOWLEDGMENT

We cannot express our gratitude to the professor for their kind patience and criticism. Without the knowledge and expertise of our professor, we would not be able to do this research. In addition, this effort cannot be achieved without the generous support of the Spring GX Scholarship. We also thank the members of our laboratory for their help with editing, their valuable feedback sessions, and their unconditional moral support. Lastly, we thank our Beyless collaborators for helping us collect data and conduct detailed discussions on this research. Finally, we have lost our innocence by not mentioning our family, especially our parents, brother, and spouses. They believed in us and maintained our spirit and motivation in this process.

REFERENCES

[1] E. Shelhamer, J. Long, and T. Darrell, "Fully convolutional networks for semantic segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 4, pp. 640–651, Apr. 2017.

[2] W. Weng and X. Zhu, "U-Net: Convolutional networks for biomedical image segmentation," *IEEE Access*, vol. 9, pp. 16591–16603, 2015.

[3] A. Paszke, A. Chaurasia, S. Kim, and E. Culurciello, "ENet: A deep neural network architecture for real-time semantic segmentation," Jun. 2016, *arXiv:1606.02147*.

[4] H. S. Choi, Y. M. Kim, and J. H. Kim, "Robust lane detection by cascaded deep learning network," in *Proc. Int. Conf. Robot Intell. Technol. Appl.* (Lecture Notes in Networks and Systems), vol. 429, 2022, pp. 303–312.

[5] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Nov. 2013, pp. 580–587.
[6] R. Girshick, "Fast R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1440–1448.
[7] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017.
[8] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 779–788.
[9] F. Scarselli, M. Gori, A. C. Tsoi, M. Hagenbuchner, and G. Monfardini, "The graph neural network model," *IEEE Trans. Neural Netw.*, vol. 20, no. 1, pp. 61–80, Jan. 2009.
[10] P. W. Battaglia, J. B. Hamrick, V. Bapst, A. Sanchez-Gonzalez, V. Zambaldi, M. Malinowski, A. Tacchetti, D. Raposo, A. Santoro, R. Faulkner, and C. Gulcehre, "Relational inductive biases, deep learning, and graph networks," Jun. 2018, *arXiv:1806.01261*.
[11] W. L. Hamilton, R. Ying, and J. Leskovec, "Inductive representation learning on large graphs," in *Proc. Adv. Neural Inf. Process. Syst.*, Jun. 2017, pp. 1025–1035.
[12] Z. Wang, T. Chen, J. Ren, W. Yu, H. Cheng, and L. Lin, "Deep reasoning with knowledge graph for social relationship understanding," in *Proc. IJCAI Int. Joint Conf. Artif. Intell.*, 2018, pp. 1021–1028.
[13] D. Zhao, Y. Guo, and Y. J. Jia, "TrafficNet: An open naturalistic driving scenario library," in *Proc. IEEE 20th Int. Conf. Intell. Transp. Syst. (ITSC)*, Oct. 2017, pp. 1–8.
[14] F. Montanari, R. German, and A. Djanatliev, "Pattern recognition for driving scenario detection in real driving data," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Oct. 2020, pp. 590–597.
[15] S. Ulbrich, T. Menzel, A. Reschka, F. Schuldt, and M. Maurer, "Defining and substantiating the terms scene, situation, and scenario for automated driving," in *Proc. IEEE 18th Int. Conf. Intell. Transp. Syst.*, Sep. 2015, pp. 982–988.
[16] R. Queiroz, T. Berger, and K. Czarnecki, "GeoScenario: An open DSL for autonomous driving scenario representation," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2019, pp. 287–294.
[17] M. Bojarski, D. D. Testa, D. Dworakowski, B. Firner, B. Flepp, P. Goyal, L. D. Jackel, M. Monfort, U. Müller, J. Zhang, X. Zhang, J. Zhao, and K. Zieba, "End to end learning for self-driving cars," Apr. 2016, *arXiv:1604.07316*.
[18] T. P. Lillicrap, J. J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, and D. Wierstra, "Continuous control with deep reinforcement learning," in *Proc. 4th Int. Conf. Learn. Represent.*, Sep. 2015, pp. 1–14.
[19] A. G. Cunningham, E. Galceran, R. M. Eustice, and E. Olson, "MPDM: Multipolicy decision-making in dynamic, uncertain environments for autonomous driving," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2015, pp. 1670–1677.
[20] M. Wang, Z. Wang, J. Talbot, J. C. Gerdes, and M. Schwager, "Game-theoretic planning for self-driving cars in multivehicle competitive scenarios," *IEEE Trans. Robot.*, vol. 37, no. 4, pp. 1313–1325, Aug. 2021.
[21] K. J. C. Fransen, J. A. W. M. van Eekelen, A. Pogromsky, M. A. A. Boon, and I. B. F. Adan, "A dynamic path planning approach for dense, large, grid-based automated guided vehicle systems," *Comput. Oper. Res.*, vol. 123, Nov. 2020, Art. no. 105046.
[22] H. Yu, K. Meier, M. Argyle, and R. W. Beard, "Cooperative path planning for target tracking in urban environments using unmanned air and ground vehicles," *IEEE/ASME Trans. Mechatronics*, vol. 20, no. 2, pp. 541–552, Apr. 2015.
[23] N. Dogru and A. Subasi, "Traffic accident detection using random forest classifier," in *Proc. 15th Learn. Technol. Conf. (LT)*, Feb. 2018, pp. 40–45.
[24] A. B. Parsa, H. Taghipour, S. Derrible, and A. Mohammadian, "Real-time accident detection: Coping with imbalanced data," *Accident Anal. Prevention*, vol. 129, pp. 202–210, Aug. 2019.
[25] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 834–848, Apr. 2018.
[26] D. Wu, M.-W. Liao, W.-T. Zhang, X.-G. Wang, X. Bai, W.-Q. Cheng, and W.-Y. Liu, "YOLOP: You only look once for panoptic driving perception," *Mach. Intell. Res.*, vol. 19, no. 6, pp. 550–562, Dec. 2022.
[27] S. Nham, J. Lee, and S. Kamijo, "DCSEG-Net: Driver's cognition based semantic segmentation by multi-level feature extraction," in *Proc. IEEE Int. Conf. Intell. Transp. Syst. (ITSC)*, Bilbao, Spain, Sep. 2023.
[28] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder–decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017.
[29] F. Karim, S. Majumdar, H. Darabi, and S. Chen, "LSTM fully convolutional networks for time series classification," *IEEE Access*, vol. 6, pp. 1662–1669, 2018.
[30] X. Chen, K. Kundu, Y. Zhu, A. G. Berneshawi, H. Ma, S. Fidler, and R. Urtasun, "3D object proposals for accurate object class detection," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 28, 2015, pp. 1–9.
[31] X. Liang, Z. Hu, H. Zhang, L. Lin, and E. P. Xing, "Symbolic graph reasoning meets convolutions," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 31, 2018, pp. 1–11.
[32] W. Zhang, Y. Zhang, L. Xu, J. Zhou, Y. Liu, M. Gu, X. Liu, and S. Yang, "Modeling IoT equipment with graph neural networks," *IEEE Access*, vol. 7, pp. 32754–32764, 2019.
[33] P. Velickovic, A. Casanova, P. Lio, G. Cucurull, A. Romero, and Y. Bengio, "Graph attention networks," in *Proc. 6th Int. Conf. Learn. Represent.*, 2017, pp. 1–12.
[34] M. Schlichtkrull, T. N. Kipf, P. Bloem, R. van den Berg, I. Titov, and M. Welling, "Modeling relational data with graph convolutional networks," in *Proc. Eur. Semantic Web Conf.* (Lecture Notes in Computer Science), vol. 10843, 2017, pp. 593–607.
[35] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: Optimal speed and accuracy of object detection," Apr. 2020, *arXiv:2004.10934*.
[36] P. Purkait, C. Zhao, and C. Zach, "SPP-Net: Deep absolute pose regression with synthetic views," Dec. 2017, *arXiv:1712.03452*.
[37] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Dec. 2016, pp. 2117–2125.
[38] F. Yu, H. Chen, X. Wang, W. Xian, Y. Chen, F. Liu, V. Madhavan, and T. Darrell, "BDD100K: A diverse driving dataset for heterogeneous multitask learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 2633–2642.
[39] Y. Li, T. Song, and X. Wu, "Robust construction of spatial–temporal scene graph considering perception failures for autonomous driving," in *Proc. IEEE Int. Conf. Intell. Transp. Syst. (ITSC)*, Bilbao, Spain, Sep. 2023.
[40] M. Teichmann, M. Weber, M. Zöllner, R. Cipolla, and R. Urtasun, "MultiNet: Real-time joint semantic reasoning for autonomous driving," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2018, pp. 1013–1020.
[41] Y. Qian, J. M. Dolan, and M. Yang, "DLT-Net: Joint detection of drivable areas, lane lines, and traffic objects," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 11, pp. 4670–4679, Nov. 2020.
[42] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 6230–6239.
[43] C. Han, Q. Zhao, S. Zhang, Y. Chen, Z. Zhang, and J. Yuan, "YOLOpv2: Better, faster, stronger for panoptic driving perception," Aug. 2022, *arXiv:2208.11434*.

SEUNGWOO NHAM (Graduate Student Member, IEEE) received the B.E. and M.E. degrees from Kyushu University, Fukuoka, Japan, in 2016 and 2018, respectively. He is currently pursuing the Ph.D. degree with the Institute of Industrial Science, The University of Tokyo. He started his career as an Associate Research Engineer with Carnavi.com, in 2018. He joined A.I.Matics, in 2020. His research interests include computer vision, point cloud processing, deep learning, and their applications in various industries and autonomous vehicles.

JINHO LEE (Graduate Student Member, IEEE) received the B.E. and M.E. degrees from Kyushu University, Fukuoka, Japan, in 2016 and 2018, respectively. He is currently pursuing the Ph.D. degree with the Institute of Industrial Science, The University of Tokyo. He was a Visiting Student Researcher with the University of California at Berkeley, in 2016. He began working with A.I.Matics., in 2018, as an Associate Research Engineer. His research interests include computer vision, localization, deep learning, and their applications to ITS and an autonomous vehicle. He was a recipient of the 69th IEEE Kyushu Section Excellent Presentation Award, in 2016, and the 69th Meeting on Image Recognition and Understanding (MIRU) Interactive Presentation Award, in 2017.

SEONGRYUL YANG received the bachelor's degree from Kyungpook National University, Daegu, South Korea, in 1994. He is currently pursuing the Ph.D. degree with Dankook University, South Korea. In 1995, he started his career as an Automotive Engineer with Hyundai Electronics' Electronics Division (merged into Hyundai Mobis). He started his own business and managed it for 14 years before joining Unitrontech. He joined Beyless, in 2023. His research interests include automotive, deep learning, autonomous vehicles, and their applications.

JIHUN KIM received the B.E. and M.E. degrees from Kyungpook National University, Daegu, South Korea, in 2013 and 2015, respectively. He started his career as a Research Engineer with Beyless, in 2019. His research interests include computer vision and deep learning, specifically focusing on tracking, object detection, and their applications to various industries and autonomous vehicles.

SHUNSUKE KAMIJO (Senior Member, IEEE) is currently an Associate Professor with The University of Tokyo. He has been engaged in video surveillance, sensor fusion, and their application to intelligent transport systems. His current research interests include autonomous vehicles, digital maps, image processing, deep learning, and portable and wearable applications. He is a Senior Member of the IEEE ITS Society. He is a member of IEICE and IATSS. He has been serving on the Board of Governors for the IEEE ITS Society, since 2015, and the Vice President for Publication, since 2022. He has many contributions to international conferences as the Vice-Chairperson of the Program Committee for the ITS World Congress Tokyo 2013, the General Co-Chair for IEEE ICVES2015 and ITSC2017, and the International Program Chair for IEEE IV2017. He also has significant contributions to international journals as the Editor-in-Chief of *IATSS Research* (Elsevier) and an Editorial Board Member of *Multimedia Tools and Applications* (Springer).