Knowledge-Guided Semantic Segmentation
Autonomous Vehicles Using Conceptual Metric Learning

Dr. R. Nagalakshmi, V. Ramesh Babu, V. Varun, A. Vishaal
asst. prof. CSE Dept., student, Dept. of CSE, student, Dept. of CSE, Karthikeyan, student,
SRMIST, Chennai, SRMIST, Chennai, SRMIST, Chennai

ABSTRACT:
Sensing is the main function of all autonomous driving and collects all the necessary information about the environment around the vehicle's movement. The decision making process uses the information needed to create strategies and make the right decision based on the situation to ensure the safety of passengers as much as possible. This article reviews recent literature on motor vehicle awareness (AVP), focusing on two main tasks: semantic segmentation and object detection. These two functions play an important role as an important part of the car navigation system. In the autonomous car, perception with point cloud semantic segmentation helps obtain a wealth of information about the surrounding road environment. Despite the massive progress of recent researches, the existing machine learning networks are still insufficient for online applications of autonomous driving due to too subdivided classes, the lack of training data, and their heavy computing load.

Keyword: Local Interpretable Model-Agnostic Explanations, adaptive dehazing, metric learning

I. INTRODUCTION
Safe navigation in adverse weather conditions is crucial for the successful deployment of autonomous vehicles. Hazy weather significantly impacts the ability of these vehicles to accurately segment their surroundings, which is essential for tasks like object detection and path planning. The successful deployment of autonomous vehicles hinges on their ability to navigate safely in diverse weather conditions. One critical task for autonomous vehicles is semantic segmentation, which involves dividing the surrounding environment into meaningful categories like lanes, pedestrians, and traffic signs. This information forms the foundation for tasks like object detection, path planning, and ultimately, safe navigation. However, hazy weather conditions pose a significant challenge to accurate semantic segmentation. Haze, a type of atmospheric scattering, reduces visibility and obscures details crucial for proper object identification.

Current deep learning-based semantic segmentation models, while achieving impressive performance under clear weather conditions, often struggle in hazy environments. These models are trained on large datasets of clear images and learn to identify objects based on specific visual features like color, shape, and texture. When presented with hazy images, these features become distorted and less distinguishable. Consequently, the models misinterpret the scene, leading to inaccurate segmentation results. This can have severe consequences, as misidentified objects or lanes can cause the autonomous vehicle to make poor decisions, jeopardizing safety.

The quest for robust and reliable autonomous vehicles hinges on their ability to navigate diverse weather conditions effectively. Semantic segmentation, a crucial task that parses the
surrounding environment into meaningful categories like lanes, pedestrians, and traffic signs, forms the bedrock for higher-level functionalities like object detection and path planning. However, hazy weather throws a significant wrench into this process. Haze, a product of atmospheric scattering, reduces visibility and obscures details critical for accurate object identification. Consequently, traditional deep learning-based segmentation models, trained on pristine datasets, falter when confronted with hazy conditions.

These models, despite achieving remarkable performance under clear skies, rely heavily on specific visual features like color, shape, and texture during object recognition. Hazy environments distort these features, making them less distinguishable. This leads to misinterpretations of the scene by the model, resulting in inaccurate segmentation. Inaccurately segmented objects or lanes can have dire consequences, potentially causing the autonomous vehicle to make poor decisions and jeopardizing safety.

Current approaches primarily tackle this challenge through data augmentation techniques. These techniques artificially introduce variations in the training data, attempting to simulate the effects of haze. However, this approach has limitations. Capturing the full spectrum of real-world hazy scenarios with sufficient nuance is a difficult task due to the sheer variability of haze intensity and atmospheric conditions. Additionally, traditional dehazing methods, while improving overall visibility, often lack the necessary finesse. These methods might over-dehaze the image, inadvertently introducing artifacts that further confuse the segmentation model.

II. RELATED WORKS

“This paper proposes a new approach for single image dehazing that leverages the median filtering property of haze-free images to improve haze estimation, especially in challenging scenarios. The proposed method, termed Joint Median Channel Prior (JMCP), incorporates both channel priors and the median filtering property to achieve superior dehazing results. The channel priors are established by analyzing the relationship between the hazy and haze-free images in different color channels. The median filtering property is exploited to refine the transmission map, a key component for haze removal. By combining these elements, JMCP effectively addresses the limitations of existing methods and achieves better haze removal accuracy, particularly in challenging cases. [1] This paper uses explainable AI for credit card fraud detection which is a new way of approach. The integration of Explainable Artificial Intelligence (XAI) into credit card fraud detection models has gained significant traction in recent years. This focus on explainability stems from the need for transparency and fairness in such high-stakes financial applications. Traditional fraud detection models, often based on complex machine learning algorithms, can be opaque in their decision-making processes. This lack of transparency can make it difficult to understand why a transaction is flagged as fraudulent, hindering user trust and potentially leading to false positives. Several research efforts have explored the application of XAI techniques to credit card fraud detection. Some studies have investigated the use of model-agnostic methods like LIME (Local Interpretable Model-Agnostic Explanations) and SHAP (SHapley Additive exPlanations) [1, 2]. These techniques provide local explanations for individual transaction classifications, highlighting the specific features that most influenced the model's decision. This allows human analysts to understand the rationale behind the model's predictions and potentially identify potential biases or weaknesses in the training data.” [2]
evaluation index, so as to improve the details of the picture. Then, the high-level semantic features of the target are learned by the AlexNet, and the RPN network is adopted for classification and regression. Moreover, based on the modeling concept of DIOU and penalty items, the interference caused by redundant anchors is eliminated to obtain a more accurate bounding box. The improved algorithm effectively enhances the tracking accuracy and robustness for haze video sequences. Our work provides a new idea for the application of object tracking algorithm to a wider range of scenarios.”[3]

“We propose a novel end-to-end adaptive enhancement dehazing network called AED-Net for single-image dehazing. It consists of gray level dehazing (GLD) based on the Res2Net encoder and a two-branch decoder, a novel region-aware modified Gamma correction (RAMGC), and a residual channel attention network (RCAN) that effectively combines high-level and low-level features. An ablation study demonstrated the effectiveness of the RAMGC module. Quantitative comparisons based on five key-metrics PNSR, SSIM, FSIM, VIF, and LPIPS, as well as qualitative analysis, were performed. The experimental results demonstrate that the proposed method outperforms state-of-the-art methods on benchmark datasets, real-world images, and the NH-Haze2 dataset of the Codalab NTIRE 2021 competition. We conclude that our novel RAMGC applied in conjunction with the two-branch treatment of haze contributes significantly to dehazing performance. We intend to explore dilated convolution with reduced parameters and verify the systems performance on other datasets for further research. Another promising area of future inquiry is combining the proposed model with the de-raining algorithm. (Sargis A. Hovhannisyan and Hayk A. Gasparyan contributed equally to this work.)”[4]

“The paper focuses on a methodology or algorithm designed to extract significant temporal patterns or sequences from time-series data in the context of Explainable AI (XAI). The paper is expected to delve into how temporal order within time-series data can be crucial for understanding and explaining the behavior of AI models, especially in dynamic environments where data evolves over time. It may introduce a novel approach or technique that not only identifies important temporal patterns but also provides explanations or insights into why these patterns are significant. Furthermore, the paper is likely to discuss the importance of interpretability in AI models, particularly in the context of time-series data analysis where decisions based on historical trends or sequential events need to be justified and understood by users. It may explore the implications of extracting temporal order for enhancing the transparency, trustworthiness, and acceptability of AI systems, especially in domains where interpretability is critical, such as finance, healthcare, or predictive maintenance.”[5]

“This paper introduces a novel methodology for enhancing the quality of nighttime images by leveraging a deep hierarchical network originally trained on daytime hazy images. Through the adoption of a deep learning-based approach, the paper is expected to utilize a hierarchical network architecture to capture and model the complex relationships within the data, a common choice for image processing tasks, due to its ability to automatically learn hierarchical representations from raw data. By training the network on daytime hazy images, the authors likely exploit the similarities between daytime and nighttime image characteristics, indicative of transfer learning, where knowledge gained from one domain is transferred to another related domain to boost performance. The focus of the paper is likely on enhancing the visibility and quality of nighttime images, which are often degraded due to factors like low light conditions and atmospheric effects such as haze. The proposed method is anticipated to effectively remove haze and other artifacts from nighttime images, resulting in clearer and more detailed visuals. The methodology presented in the paper could have significant implications in various fields where nighttime imagery is essential, including surveillance, security, and environmental monitoring, enabling better
decision-making and analysis in these domains. Overall, the paper is expected to contribute to the advancement of image processing techniques tailored specifically for nighttime conditions, demonstrating the efficacy of deep learning and cross-domain training strategies in enhancing image quality and visibility in challenging low-light environments.”[6]

“The paper focuses on methodology for improving the clarity and quality of high-resolution multispectral remote sensing images through a data-driven approach. This approach is expected to leverage advanced techniques in image processing and machine learning to effectively remove haze and atmospheric interference from remote sensing imagery. By employing a data-driven approach, the paper likely aims to analyze and learn from a large dataset of multispectral remote sensing images, enabling the development of robust dehazing algorithms that can handle various environmental conditions and imaging scenarios. Furthermore, the focus on high-resolution multispectral imagery indicates the importance of preserving spectral information and fine details during the dehazing process, which is crucial for accurate analysis and interpretation in applications such as environmental monitoring, land use classification, and disaster management. The proposed methodology is anticipated to contribute to the advancement of remote sensing technology by enhancing the visibility and interpretability of multispectral imagery, ultimately facilitating more accurate and insightful analysis of Earth’s surface and environmental phenomena. Overall, the paper is expected to contribute to the advancement of hardware-accelerated image processing techniques, particularly in the domain of image dehazing, by providing a dedicated VLSI architecture optimized for FPGA implementation.”[8]

“This paper presents a novel VLSI (Very Large Scale Integration) architecture designed for implementing a saturation-based image dehazing algorithm on FPGA (Field-Programmable Gate Array) platforms. The focus of the paper is expected to be on hardware-efficient design methodologies tailored specifically for image dehazing algorithms, aiming to accelerate the processing speed and efficiency of dehazing operations. By leveraging FPGA technology, the proposed architecture is anticipated to provide real-time or near-real-time performance for image dehazing tasks, making it suitable for applications requiring low-latency processing such as surveillance, autonomous vehicles, and medical imaging. Additionally, the use of a saturation-based dehazing algorithm suggests an emphasis on simplicity and computational efficiency, making it well-suited for implementation on hardware platforms with limited resources. The paper is likely to discuss the architectural design considerations, optimization techniques, and performance evaluation results of the proposed VLSI architecture, highlighting its advantages over software-based implementations in terms of speed, power efficiency, and scalability. Overall, the paper is expected to contribute to the advancement of hardware-accelerated image processing techniques, particularly in the domain of image dehazing, by providing a dedicated VLSI architecture optimized for FPGA implementation.”[7]

“The paper explores the application of explainable AI (XAI) methods to sequence autonomous network functions effectively. The focus is expected to be on developing transparent and interpretable AI models or algorithms that can autonomously sequence network functions in a manner that is understandable and justifiable to human operators or stakeholders. This approach aims to enhance the trustworthiness and reliability of autonomous networks by providing explanations for the decisions made during the sequencing process. The paper may discuss various XAI techniques such as decision trees, rule-based systems, or attention mechanisms, which can offer insights into the rationale behind the chosen sequences of network functions. Furthermore, the research may address the challenges of integrating XAI methods into autonomous network systems, including scalability, real-time performance, and the trade-
off between transparency and complexity. By elucidating the sequencing of autonomous network functions through XAI methods, the paper is expected to contribute to the development of more transparent and accountable autonomous network architectures, thereby facilitating their adoption in critical infrastructure, telecommunications, and other industries where network reliability and performance are paramount. Overall, the paper is likely to pave the way for safer and more explainable autonomous network operations, fostering trust and confidence among users and stakeholders.”[9]

“In this paper, we propose a novel haze removal algorithm using a multiple scattering model. Unlike the most of the existing approaches are based on the single scattering model, or spatially invariant blur kernel, we proposed a spatially variant atmospheric point spread function with superpixel algorithm. Moreover, the generalized normal distribution is employed to model the physical blur kernel caused by multiple scattering, atmospheric point spread function. We define the blur kernel of each region with three different prior for characteristics of the road scenes: the angle norm factor, gradient value, and the modified angle norm factor. To prevent artifacts from edges and remove noise, the total variation regularization is adopted.”[10]

The paper titled "Building Confidence and Acceptance of AI-based Decision Support Systems - Explainable and Liable AI 2020" likely explores strategies and methodologies aimed at enhancing the trust, confidence, and acceptance of AI-based decision support systems through the incorporation of explainable and liable AI principles. In particular, the paper is expected to address the growing need for transparency and accountability in AI systems, especially those that play a crucial role in supporting decision-making processes in various domains. The focus is likely on developing AI models and algorithms that not only provide accurate predictions or recommendations but also offer explanations for their outputs, allowing users to understand the rationale behind the decisions made by the system. Additionally, the paper may discuss the importance of ensuring the liability of AI systems, particularly in contexts where decisions have significant consequences or impact on individuals or society as a whole. This may involve implementing mechanisms for tracing and auditing AI decisions, as well as establishing frameworks for assigning responsibility in case of errors or biases. By emphasizing the principles of explainability and liability in AI-based decision support systems, the paper is expected to contribute to building trust and acceptance among users, stakeholders, and regulatory authorities, ultimately facilitating the widespread adoption and integration of AI technologies in decision-making processes across various sectors. Overall, the paper is likely to provide valuable insights and guidelines for developing responsible and trustworthy AI systems that prioritize transparency, accountability, and user-centric design.”[11]

Existing System:

“Haze-level discriminators are crucial for autonomous vehicles to handle segmentation tasks successfully in hazy and foggy outdoor environments. Deep learning (DL) networks trained to segment clear images exhibit more false positives and fail to recognize the pixel patterns for the class categories when faced with hazy images. To address this issue, the existing system authors propose a novel dehazing scheme called Adaptive Dehazing (AD) passing them to the DL for segmentation. We define various thresholds to classify hazy images into four categories: heavy, moderate, slight, and clear. Additionally, the existing system authors implement a hazy image generator to create hazy synthetic images that replicate actual hazy road scene conditions for testing the existing algorithm. Moreover, the existing system authors use the Explainable Artificial Intelligence (XAI) method to understand the feature selected by different layers in the network before and after applying the AD and their contribution to obtaining a final segmented output. Extensive experiments demonstrate both quantitatively and qualitatively the performance
Improvement in the segmentation task by achieving an optimal balance between Intersection over Union (IoU) and pixel accuracy (PA) metrics of the segmented categories. The improvement in IoU metrics shows that the AD scheme significantly outperforms the previous state-of-the-art dehazing methods. The introduction of autonomous vehicles (AVs) into the heavily congested roads of various nations has led to a fundamental change in how human transportation is conceptualized and implemented. As part of this significant transition, the responsibility for vehicle control is currently shifting from human drivers to onboard computer systems, which require precise road scene segmentation techniques that are robust to varying weather conditions and ensure safety for future road transportation. This paper presents a novel adaptive dehazing scheme that adjusts to the level of haze. Additionally, the study addresses the lack of haze-level discrimination and segmentation evaluation methods, which fills a significant gap in the literature. The proposed system uses a convolutional structure with a stride of 4 and a kernel size of 11 * 11 with LayerNorm (LN). The input image size is converted from B*C*H*W to B*C*H/4*W/4, where B refers to the batch size; C and C are the number of channels, at 3 and 32, respectively, and H and W are the height and width of the input image, respectively. In contrast to directly cutting the image, the patch partition technique can extract the features of the image once and can be understood as preserving the connection information between different regions of the image to a certain extent. The design of the decoder process to avoid the introduction of a large amount of computation and a complex structure. To improve the segmentation detail performance of the Seg-Road model, convolutional layer was introduced in the decoder. Further, the convolutional layer layer was used to calculate the feature maps of the stitched channels and obtain the final output.

Proposed System:

The network adopts an encoder-decoder structure. “The encoding stage is used for feature extraction, and the decoding stage is used to restore the original size and output semantic segmentation results. Besides, skip connections are added to the network so that the recovered feature maps can have more low-scale features to obtain better semantic segmentation results. We use an asymmetric encoder-decoder structure in the network, where the PointTensor and Cylinder voxels are fused once in the encoding stage and twice in the decoding stage. The PointTensor needs more information from the Cylinder voxels in the decoding stage to ensure that the recovered feature map has more details to make the final output semantic segmentation result more accurate. The proposed system uses a convolutional structure with a stride of 4 and a kernel size of 11 * 11 with LayerNorm (LN). The input image size is converted from B*C*H*W to B*C*H/4*W/4, where B refers to the batch size; C and C are the number of channels, at 3 and 32, respectively, and H and W are the height and width of the input image, respectively. In contrast to directly cutting the image, the patch partition technique can extract the features of the image once and can be understood as preserving the connection information between different regions of the image to a certain extent. The design of the decoder process to avoid the introduction of a large amount of computation and a complex structure. To improve the segmentation detail performance of the Seg-Road model, convolutional layer was introduced in the decoder. Further, the convolutional layer layer was used to calculate the feature maps of the stitched channels and obtain the final output.” (doaj.org, n.d.)
III. Methodology

This part “primary work is preparing datasets. In this paper, the pictures of corn plants are self-built (see the experiment and discussion section for details), and the real-value pictures are manually annotated. To simplify the calculation and operation, the size of the picture is adjusted uniformly, and the RGB image is converted into an HSV image. The geometric transformation methods mainly include flip, rotation, cropping, scaling, translation, etc. To prevent overfitting, pixel transformation methods such as adding salt and pepper noise, Gaussian blur, etc., can also be used. Finally, it is divided into training, verification, and test subsets according to a specific ratio. Just like UNet, whose training data is a set of 30 images (512 * 512 pixels) from serial section transmission electron microscopy of the Drosophila first instar larva ventral nerve cord (VNC), we only needed a tiny raw dataset, and the images and videos collected in our field were separated into 30 images of individual corn, which was very meaningful for alleviating the burden of data collection. The image annotation tool Labelme created the semantic annotation map corresponding to the plant image, using the function named cvtColor() of the open computer vision library OpenCV2 to convert plant images into HSV format images. The RGB and HSV images shared the same semantic annotation map. We used the ImageDataGenerator class of the artificial neural network library Keras to achieve image enhancement (rotation, movement, scaling, etc.). It should be noted that the semantic annotation map and the raw image are converted synchronously. Finally, we brought the amount of the total dataset to 4000 * 3, and each type of image (RGB, HSV, and Mark) was divided into a training set, a verification set, and a test set in a ratio of 7:2:1.” (ncbi.nlm.nih.gov, n.d.)

FEATURE SCALING

“Feature Scaling is to keep the channel of feature fusion the same. For the feature layer that needs to be upsampled, first use 1 * 1 convolution to adjust the number of channels to be consistent with the target layer, and then use interpolation to increase the resolution and adjust the size. For the 1/2 scale downsampling layer, a convolution of size 3 * 3 with stride 2 is used. For the 1/4 scale downsampling layer, it is necessary to add a maximum pooling layer with a stride of 2 to the convolution with a size of 3 * 3 and a stride of 2.” (mdpi.com, n.d.) “Objects in images vary in size and shape, and features extracted at a single scale might not effectively capture all relevant information on objects shown in images. Multiscale feature extraction involves using multiple convolutional kernels or pooling layers with different scales to capture the features of objects at various scales. This technique enables a classifier to identify objects more accurately, especially when images show objects of various sizes. Multiscale feature extraction is an effective technique for enhancing the accuracy and robustness of classifiers and provides excellent performance for the recognition of objects of varying sizes within images. CNNs use the receptive field to perceive features at different scales. If the receptive field is too small, the network can only capture local features; if the receptive field is too large, excessive noise might be captured.” (www2.mdpi.com, n.d.)

CONCEPTUAL MATRIC LEARNING NETWORK

In order to associate these two branches, we employ a domain transform structure to combine the coarse semantic segmentation method with the edge detection in a parallel fashion. Therefore, our model can refine the coarse semantic segmentation guided by the edge information, and also be jointly trained end-to-end. The domain transform we employ requires two different inputs. One is the original input signal X, which corresponds to the coarse semantic segmentation results in our model. The other is a positive domain transform density signal d, which relates to the edge prediction map. The output of the domain transform is a filtered signal Y, which corresponds to the final refined semantic segmentation. We will introduce the recursive formula of domain transform to explain how the
filtered signal $Y$ is obtained. At present, the loss functions commonly used include Cross Entropy loss, Mean Square Error loss or Mean Absolute Error loss. The Cross Entropy describes the probability of classification correlation between the two image distributions, which is simple and accurate. In addition, semantic segmentation itself can be regarded as a multi classification problem, that is, each target type can only be represented by one category value, and so Cross Entropy loss function is frequently used in semantic segmentation tasks.

EVALUATION METRICS

“Semantic segmentation metrics play an essential role in evaluating performance techniques. Different semantic segmentation assessment criteria may produce disparate results because it is unclear how to define successful performance segmentation. Pixel accuracy, mean intersection over union, and representing per-class accuracy are three of the most frequently used measures. For all of them, let $n$ be the number of class, pixels predicted to belong to the classes, $j$. In addition, let $k_i$ be the total pixel number belonging to class $i$. If we assume to have a $T$ total number of classes,” (mdpi.com, n.d.) then:

1. Pixel accuracy: can be computed as

$$acc = \frac{\sum n_j}{\sum_i n_j}$$

2. Mean intersection over union: can be computed as

$$mIoU = \frac{1}{T} \frac{\sum_i n_{ij}}{\sum_i k_j}$$

$$mIoU = \frac{1}{T} \frac{\sum_i n_{ij}}{\sum_i k_j + \sum_j n_{i,j} - n_{ij}}$$

3. Mean per class accuracy: can be computed as:

The hazy image is fed into the module. The image might undergo preprocessing as mentioned above (represented by Encoder (P) in your diagram). The core functionality of the dehazing module lies here. It estimates a transmission map that describes how much light is scattered by haze particles at each pixel location. Additionally, it estimates the atmospheric light, which is the color of the light source that gets scattered by the haze. Using the estimated transmission map and atmospheric light, the dehazing algorithm recovers the original scene radiance or the dehazed image.
RESULT

“In view of the results, although PointNet allows a segmentation of the environment with good overall accuracy, in order to segment small specific elements, there are already other methodologies that achieve better success rates without the need for such costly training. Practitioners should evaluate whether to include classes that contain a relative low number of points that may be considered as noisy assets. “ (mdpi.com, n.d.)

CONCLUSION:

In this work, a methodology for semantic segmentation of continuous elements “conforming road environments in urban point clouds has been presented. A methodology has been implemented to segment the point cloud into sections along the road, which allows the methodology scalability regardless of road length. Semantic segmentation has been performed directly on the point cloud using PointNet. The intensity, the return number and the total number of returns have been used as false color information.

In view of the results, although PointNet allows a segmentation of the environment with good overall accuracy, in order to segment small specific elements, there are already other methodologies that achieve better success rates without the need for such costly training. Practitioners should evaluate whether to include classes that contain a relative low number of points that may be considered as noisy assets. In this work, we overviewed the current approaches for multimodal road-scenes segmentation, with particular attention to the imaging modalities and datasets used. Several different approaches have been discussed and compared, showing how the combination of multiple inputs allows for improving the performance with respect to each modality when used alone. Even if there is a variety of different solutions, it is possible to notice a quite common design strategy based on having one network branch for each modality and some additional modules moving the information “ (ncbi.nlm.nih.gov, n.d.) across them or merging the extracted features.
REFERENCE

1. Joint Median Channel Prior and Edge Enhancement for Single Image Dehazing 2023
2. Explainable AI supported Evaluation and Comparison on Credit Card Fraud Detection Models 2023
3. A Deep Learning Approach to Single Image Dehazing Inspired by Euler Numerical Scheme 2023
4. Explainable AI (XAI) for AI-Acceptability: The Coming Age of Digital Management 5.0 2023
5. Extraction of Important Temporal Order for eXplainable AI on Time-series data 2023
6. Night-time image dehazing using deep hierarchical network trained on day-time hazy images 2023
8. VLSI Architecture of Saturation Based Image Dehazing Algorithm and its FPGA Implementation 2022
9. Sequencing of Autonomous Network functions using eXplainable AI Methods 2022
10. Single Image Dehazing Using Wavelet-Based Haze-Lines and Denoising 2021
11. Build confidence and acceptance of AI-based decision support systems - Explainable and liable AI 2020