Introduction：

As autonomous driving technology continues to advance, a deep understanding of the surrounding environment becomes particularly critical in order to achieve safe and accurate navigation of vehicles in complex natural environments. Autonomous vehicles should behave differently when driving in different landscapes such as gravel, asphalt, sand or mud, which requires the vehicle to be able to accurately identify various scenes and objects in its path. This task is particularly challenging in natural environments, where scenes often exhibit highly irregular and unstructured features.

In this context, image semantic segmentation plays a crucial role as a technique to understand key information in complex environments. The goal of semantic segmentation is to classify the images captured by the vehicle camera at the pixel level, that is, to assign a specific category label to each pixel in the image, which helps the autonomous driving system to accurately identify and process various conditions ahead.

This study will use a multimodal dataset called WildScenes, which includes a total of 9,306 images from five sequences captured by ordinary video cameras in Venman National Park and Karawatha Forest Park in Brisbane, Australia. Each image has a resolution of 2016 x 1512 pixels. These images have been manually annotated, giving us accurate ground-reality data, which is a valuable resource for detailed analysis and model training. In this project, we will focus on using these 2D images and ignore the 3D point cloud portion of the dataset.

This report will introduce in detail the method of image semantic segmentation using U-net and Deeplabv3 models. U-net model has proved its effectiveness in medical image segmentation because of its unique symmetrical structure and excellent detail capturing ability. The Deeplabv3 model, an important milestone in the field of deep learning, optimizes segmentation accuracy and processing speed with its efficient spatial pyramid pooling and atrous convolution technology. Through the experiment and analysis of these two models, this report aims to demonstrate their application effects and potential in dealing with autonomous driving scenarios.

In the following chapters, we will discuss in detail the working principle, experimental process and results of these models, as well as their performance in autonomous driving image segmentation tasks, and discuss the results and method performance to provide technical support and theoretical basis for the navigation of autonomous vehicles in natural environments.

Literature review:

* 1. Background and Significance

The development of autonomous driving systems is gradually changing the way we transport, but for these systems to operate effectively in their natural environment, they must be able to address a unique and complex set of challenges. Natural environments are significantly more unpredictable and diverse than relatively structured urban environments. These environments include not only diverse surfaces such as gravel, sand, and mud, but also natural obstacles such as bodies of water and dense vegetation, each of which places different demands on autonomous vehicles' perception systems and navigational decision-making capabilities.

In such environments, the safety and efficiency of autonomous vehicles depend on high-precision environmental perception. The image captured by the vehicle camera can be segmented by careful semantic segmentation, which can not only distinguish which areas are drivable and which are not, but also identify various potential obstacles and hazards. For example, accurately identifying and classifying sandy and muddy areas in images is critical to preventing vehicles from getting stuck or coasting.

In addition, semantic segmentation technology is equally important for achieving dynamic interaction between vehicles and the environment. Driving in a natural environment, the vehicle may need to respond in real time to sudden natural events, such as the sudden appearance of animals or temporary puddles. By assigning precise category labels to each pixel, semantic segmentation not only enhances the autonomous driving system's understanding of the environment, but also greatly improves its ability to respond to external changes, thus ensuring both safe and efficient navigation.

Therefore, exploring and optimizing semantic segmentation technology for use in natural environments is not only the key to improving the performance of autonomous vehicles, but also an important step to promote the commercialization and popularization of autonomous driving technology. By continuously improving these technologies, we are able to provide more powerful tools for autonomous vehicles to cope with changing natural environments and achieve true global autonomy.

* 1. Correlative Research and Analysis

In the field of autonomous driving, semantic segmentation research is constantly advancing with the aim of improving the navigation ability and decision-making efficiency of the system in various environments. In addition to the previously mentioned U-net and Deeplabv3 models, there are several other models that demonstrate excellent performance in achieving efficient and accurate environment awareness:

* **FCN (Full Convolutional Network)：** FCN (Full Convolutional Network) : By converting standard deep networks to full convolutional form, FCN enables end-to-end pixel-level prediction and makes a major breakthrough in semantic segmentation. Jonathan Long(2015) first proposed the FCN model in Fully Convolutional Networks for Semantic Segmentation. Liang-Chieh Chen (2017) compared the innovations of FCN and DeepLab model in using atrous convolution to improve semantic segmentation. (Long et al., 2015)
* **SegNet：** Specifically targeted at video surveillance and road scenarios, SegNet effectively recovers image details, especially border areas, through its unique codec structure. Alex Kendall (2017) proposed Bayesian SegNet, which enables SegNet to provide a confidence estimate of its prediction when processing image segmentation tasks. This is particularly important for improving the reliability of automatic driving. (Kendall et al., 2015)
* **SAM (Spatial attention model)：** Jun Fu (2019) et al. proposed the use of spatial and channel attention mechanisms to improve the accuracy of scene segmentation. By introducing spatial attention mechanisms, SAM model improves the recognition ability of important features in the scene, which is crucial for identifying key obstacles and road conditions in automatic driving. (Fu et al., 2020)

In recent years, with the development of technology, researchers are also exploring combining the characteristics of multiple models to solve specific problems in autonomous driving. For example, some research has focused on combining the precise detail recovery capability of U-net with the efficient context understanding capability of Deeplabv3, aiming to create a model that can process large images quickly while maintaining high segmentation accuracy.

* 1. Problems that Exist and Need to be Solved

Despite the progress made in this area, several challenges remain in the field of semantic segmentation for autonomous driving:

* High variability of environmental conditions: Natural scenes have less structure and more variation, making it difficult to achieve consistent segmentation accuracy.
* Real-time processing requirements: For autonomous driving, segmentation algorithms must be accurate and fast enough to support real-time decision making. Many deep learning models, while accurate, struggle to meet the speed requirements of real-time processing. (Muhammad et al., 2022)
* Limited annotation datasets: Most advanced models require large amounts of annotation data, which can be expensive and time-consuming to produce, especially for uncommon or highly specific environments.
* Adaptability and ability to generalize: Many models perform well under training conditions, but struggle to generalize to new or slightly different environments without extensive retraining or fine-tuning.