Semantic Segmentation for Self-driving Cars

Honar Mohammed Taha[1], Asst. Prof. Dr. Gullanar M Hadi[1]

[1]Software & Informatics Engineering Department (SIE), College of Engineering,
Salahaddin University-Erbil, Kurdistan-Iraq.

***Abstract*— One important computer vision challenge for increasing the precision and effectiveness of vehicle operations in autonomous driving scenarios is semantic segmentation for self-driving automobiles. The pixel-by-pixel assignment to distinct item categories is a crucial aspect of creating a thorough cognitive illustration of the scene. The paper offers a full overview and detailed examination of advanced segmentation of semantic image techniques based on deep learning, intended particularly for semantic segmentation in situations including autonomous driving. Usually, autonomous cars come with a list of acquisition devices so they may do a thorough scan and utilize their complementing features. A complete dataset comparison, spanning from the earliest to the most recent ones examined in this work, is provided to wrap up. In the article, recent convolutional neural network (CNN) architectures for semantic segmentation—which are fully convolutional networks—are studied first. The other two models are temporal and context-aware.**

***Index Terms*— self-driving cars, Semantic segmentation approaches, deep learning methodology.**

I. INTRODUCTION

Semantic segmentation, a key technique in computer vision research, was initially presented during the 1970s and tries to classify all of the scene's pixels and points into several regions with separate semantic categories (Zhang et al., 2019). The use of neural networks has made tremendous progress in recent deep-learning attempts to tackle semantic segmentation. Academics did not begin to consider neural networks until the 1990s, even though they had been around since the 1940s (Ruichek and Lateef, 2019). Deep learning's amazing discovery has had a big impact on semantic segmentation methods, improving their accuracy performance. This potential advancement has attracted interest from a wide range of scientific and technological sectors that require sophisticated computer vision capabilities. This is useful in autonomous driving, where autonomous vehicles need to comprehend everything around them, including other vehicles, pedestrians (those on foot), traffic lanes, traffic lights, and traffic signals. Because deep neural networks are so good at detecting and classifying information, deep learning-based semantic segmentation is crucial to achieving this goal (Papadeas et al., 2021).

Most of the early algorithms developed for semantic segmentation used input from an only RGB camera. But still, further development and generalization to other modalities are needed for multi-sensor onboard self-driving cars. A greater understanding of the environment is possible by merging the many sensor data streams (RGB, LiDAR, RADAR, stereo setups, etc.). (Rizzoli, Barbato and Zanuttigh, 2022). The Fully Convolutional Network (FCN) model is the first method to demonstrate the deep learning potential of this problem. It is an end-to-end convolutional model that uses an encoder, also known as a contraction segment, and a decoder, also known as an expansion segment. In the first case, the expansion block upsamples the mapped low-resolution feature representation from the input. The encoder, also known as the backbone, is a pre-trained image classification network that is frequently used as a feature extractor. ResNet, VGG, and the lighter MobileNet are well-liked options among these networks. Human mistake is the primary cause of As per the National Highway Safety Administration, 94 percent of highway accidents (Rizzoli, Barbato, and Zanuttigh, 2022). The development of Automated Driving Systems (ADS) aims to lessen driving-related stress, limit emissions, prevent accidents, and transport mobility persons. By 2050, the yearly social benefits of ADSs might be close to $800 billion if extensive deployment is accomplished. These advantages would come from reduced traffic, fewer accidents on the roads, lower energy use, and higher productivity (Yurtsever et al., 2020). The car's control system is based on information that has been processed from multiple sensors that are mounted above it. This allows the vehicle to see its environment, comprehend traffic situations, and control its motions (Muhammad et al., 2022). It consequently operates the senses of hearing and vision.

This work looks at deep learning techniques for Semantic segmentation in self-driving cars and explains them. This serves as an outline for the remainder of the paper. Semantic segmentation techniques and methodologies used are classified in Section 4. All of the sensors that are utilized in the car to prevent collisions and provide robustness are examined and listed in Section 5 of related works on multimodal perception in autonomous driving with the difficulties encountered in segmenting semantics for self-driving automobiles. Afterwards, relevant datasets are listed with their performance ranging from the oldest to the newest. Multimodal datasets are examined, contrasted, and evaluated in Section 6 according to their unique attributes, advantages and disadvantages, and

DOI:http://doi.org/10.24086/cocos2024/paper.1537

underlying frameworks. The study's conclusion is included in Section 7 and few of the trends for future research are disclosed in Section 8.

Preliminary data on profound multimodal perception in self-driving cars is presented here. We commence by providing a concise overview of widely used sensing methods used by automobile sensors, and a range of experimental and study cars.

After that, semantic segmentation and deep object detection are presented. We focus here on image-based approaches since Deep learning has been utilized largely for image-based inputs. In Section 3, We'll proceed through further techniques for handling RADAR and LiDAR data.

II. SEMANTIC SEGMENTATION METHODOLOGY

Three subcategories of semantic segmentation exist temporal models, context-aware models, and fully convolutional networks. The initial category focuses on deep learning-based semantic segmentation, whereas the second one uses fully convolutional networks to leverage temporal information and context knowledge. (Siam et al., 2018)

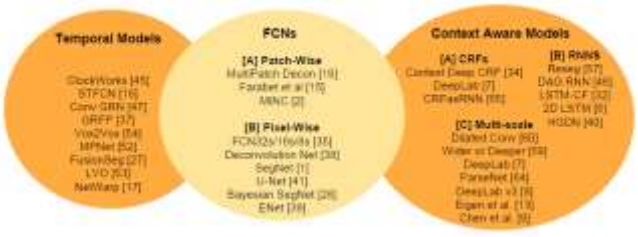

Figure 1. (Taxonomy of Semantic segmentation techniques (Siam et al., 2018))

A. FCN Fully Convolutional Networks:

The use of convolutional neural networks for semantic segmentation has progressed from patch-wise training to multi-patch training approaches that employ sliding window classification, Laplacian pyramids, and 3-stage networks. End-to-end pixel-by-pixel categorization is the main goal of deep semantic segmentation. Many techniques, such as SegNet, Bayesian SegNet, fully convolutional networks, SkipNet, and deeper decoder networks, have been proposed. According to Siam et al. (2018), these techniques make use of heatmaps, transposed convolution, layered transposed convolution, higher resolution feature maps, unpooling layers, and dropout during inference to account for prediction uncertainty. The U-Net and SkipNet8s architectures are shown in Figure 2

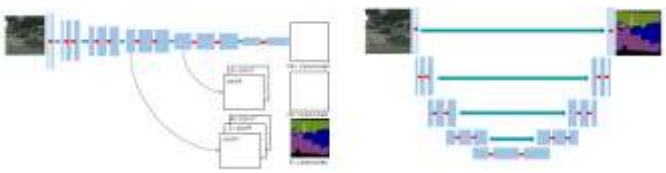

Figure 2. (form (Siam et al., 2018), on the left SkipNet(Long, Shelhamer and Darrell, no date) and on the right U-Net)

B. Context Aware Models:

Segmentation accuracy has been improved by adding context to fully convolutional networks. Recurrent neural networks, conditional random fields, and multi-scale support are some of the techniques (Siam et al., 2018). Figure 4 shows that merging multi-scale support may be done in several ways using dilated convolutions and spatial pyramid pooling.

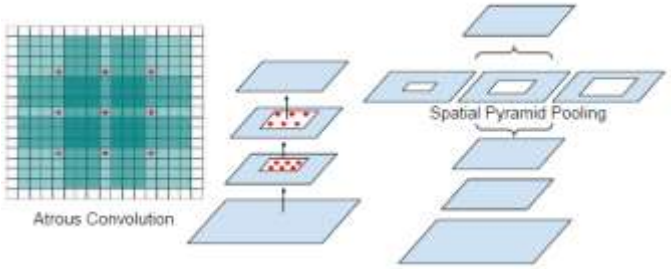

Figure 3. (As shown in (Siam et al., 2018), spatial pyramid pooling and Atrous convolution provide multi-scale assistance.)

C. Temporal Models:

Timing information has been added to recent work on video semantic segmentation using 3D convolutional networks and clockwork networks. Recurrent neural networks, on the other hand, can overcome the shortcomings of 3D convolutional networks in capturing long-term dependencies (Siam et al., 2018).

III. MAIN BODY

A. Acquisition and Multi-modal Sense:

An autonomous driving system's ability to communicate with its central perception system and exchange data with other systems depends on the infrastructure and acquisition devices chosen. Various camera, LiDAR, RADAR, GPS, and IMU configurations have been demonstrated over the years. In this part, the most often utilized sensors, their locations, and the subsequent work procedures required to turn the supplied data into a machine-readable format are outlined. Figure 4 shows a sensor configuration example. The vehicle shown in the photo was used in the production. The work's authors explain the decision to make it resemble real driverless vehicles. (Rizzoli, Barbato and Zanuttigh, 2022).

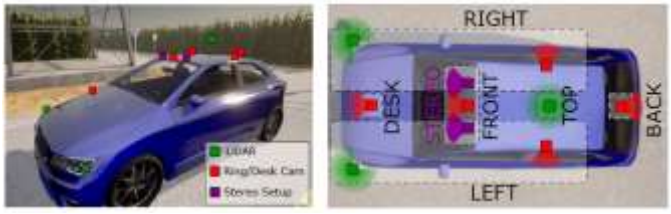

Figure 4. (It's taken from (Rizzoli, Barbato and Zanuttigh, 2022))

Many onboard senses are used by the most recent automatic driving systems (ADS). For most applications, significant sensor redundancy is essential. tasks to evaluate resilience and dependability. Five categories could be used to categorize hardware components: actuators, communication arrays, computation in units, proprioceptive sensors for automotive

DOI:http://doi.org/10.24086/cocos2024/paper.1537

status monitoring activities, and perceptual exteroceptive sensors. Extrasensory perception (EST) sensors are primarily used for detecting objects in the environment, which includes both static and dynamic objects including buildings, drivable zones, and pedestrian crossings. Lidar, camera, and the two most popular forms of sensors for this kind of job are ultrasonic and radar (Yurtsever et al., 2020).

i. RGB Cameras:
Even though conventional color cameras are often used in setups, it is often required to combine many cameras due to their low cost to get a 360° Field-of-View and improved scene interpretation. There are some restrictions with these cameras, though. For example, they cannot measure distance and are impacted by sunlight and weather. integrating color cameras with additional equipment, notably sensors, is a promising goal, particularly for low-light circumstances (Rizzoli, Barbato and Zanuttigh, 2022).

ii. LiDAR (Light Detection and Ranging):
LiDAR makes advantage of the 3D depth information collected by laser beam reflections, which allows it to classify objects more precisely than visual cameras. However, they struggle with distant things and fine textures. Precise velocity and object information is provided by FMCW LiDAR and current flash LiDAR (Feng et al., 2021) (Rizzoli, Barbato and Zanuttigh, 2022).

iii. RADAR (Radio Detection and Ranging):
While they rarely combine RGB and depth data, RADAR sensors are capable of measuring both depth and distance. They can withstand variations in weather and illumination, but their low resolution makes it difficult for them to understand semantics. Certain research indicates that Semantic segmentation setups can employ RADARs, which can also be utilized to automatically identify RADAR samples. (Feng et al., 2021) (Rizzoli, Barbato and Zanuttigh, 2022).

iv. ToF (Time of Flight):
To calculate the distance between scene points, a Time of Flight (ToF) camera monitors the light's round-trip time. Direct (dToF) and indirect (iToF) sensors are the two types that it belongs to. Despite having exceptional spatial resolution, iToF sensors have a limited range of fewer than 30 meters, making them less suitable for autonomous driving. LiDARs may capture depth information using direct (dToF) sensors because they coordinate the light pulse's arrival to and departure from the sensor (Rizzoli, Barbato and Zanuttigh, 2022).

v. HD maps with GNSS:
Systems that provide precise 3D item placements using a global satellite system and a receiver are referred to as GNSS and HD-Maps systems. Galileo, GPS, and GLONASS are GNSS examples. GPS and HD Maps were first used in car driver assistance systems, but they are now integrated to give consumers of self-driving cars course planning and ego-vehicle localization (Rizzoli, Barbato, and Zanuttigh, 2022).

vi. IMU (Inertial Measurement Units):
Dynamic driving control systems in cars have been using proprioceptive sensors, such as odometers and inertial measurement units (IMUs), to record rotational rates and accelerations for accurate localization in autonomous driving since the 1980s (Rizzoli, Barbato and Zanuttigh, 2022).

B. Sensors Setup:
Numerous sensors have been used in autonomous driving tests. In addition to Daimler installing cameras on cars, BMW testing autonomous driving in Munich since 2011, Google testing autonomous cars in more than 20 US locations, and Tartan Racing Team's Boss winning the Urban Challenge of DARPA in 2007 are all depicted in Figure 5.(Feng et al., 2021).

Figure 5. (a) DARPA Boss auto-driving car, (b) auto-drive car by Waymo (Feng et al., 2021))

C. Challenges:
1- Computational Restrictions in Systems Embedded:
The recommended approach for the Nvidia Tegra X1 car platform achieves 3 frames per second with somewhat more precision, although this is insufficient for highway driving. When the resolution is reduced to VGA (640x480), it gets close to 10 frames per second, but accuracy suffers and little objects are missed. Using 4X, complete surround view sensing requires a minimum of 4 cameras (Siam et al., 2017).

2- Annotated datasets with a high volume are needed:
The enormous Imagenet dataset made deep learning's real potential visible. Compared to Imagenet, semantic segmentation requires a much larger dataset and has much more functional complexity. It typically takes an hour or more to tag one photo using semantic segmentation annotation. the classifier's capacity to use bootstrapping, temporal propagation, and other cues like LIDAR (Siam et al., 2017).

3- Challenges in Learning:
Class disparity: takes place in the representation of significant items, like pedestrians, leading to a bias towards ignoring smaller ones. This can be resolved by employing mask predictions on bounding boxes or weighting algorithms in the error function.
i.Unobserved entities: Unseen items are not handled by the Soft-max classifier because of probability one normalization. Measuring output classification uncertainty, akin to that of a Bayesian Segnet, can be used to address covering rare objects during the training phase.
ii. Complexity of the result: Semantic segmentation generates complex contours in high-textured images, while direct classification is required for simpler object

DOI:http://doi.org/10.24086/cocos2024/paper.1537

representations, which are required for post-processing modules like mapping and maneuvering.

iii. recovering specific items: By pixel Semantic segmentation may be useful for tracking systems that monitor certain things, like pedestrians, as it produces segments of the same object rather than distinct objects.

IV. DISCUSSION & COMPARION

A. Multimodal dataset

A significant issue facing deep learning-based systems is the enormous amount of labeled data that must be produced for improvement. This is demonstrated in the domains of fake or real-world dataset creation and data categorization for deep learning model training. One critical task for autonomous driving is semantic segmentation, which has data availability issues. The intricacy of the task directly affects the amount of time and money needed to collect datasets for semantic segmentation. This study emphasizes current issues and limitations in semantic segmentation using datasets. Seldom are large-scale semantic segmentation datasets for scenarios involving autonomous driving, and even fewer consider the multimodal nature of automotive sensors. The most often used driving datasets for semantic segmentation tasks are categorized according to modalities, tasks, and data variability. (Rizzoli, Barbato, and Zanuttigh, 2022).

Shorthand notation is used to compare multi-modal datasets; it includes type, camera, daytime, location, state, labeled views, and variability. The table (Rizzoli, Barbato, and Zanuttigh, 2022) shows the scenarios that are part of each dataset.

Metadata				Sensors					Diversity				Labels			Size	
Name	Created	Update	Type	Cameras	LIDARs	Stereo	GT Depths	RADARs	IMU	Daytime	Season	Location	Weather	Env. Control	Sem. Seg.	Bboxes	Opt. Flow	Sequences	Labeled Samples
KITTI [5,45–47]	2012	2015	R	2G/2C	1	2	-	-	+	-	-	-	-	-	+	-	11(h)	200
Cityscapes [69]	2016	2016	R	2C	-	1	-	-	+	-	-	27C †	-	-	+	-	-	5000
Lost and found [69]	2016	2016	R	2C	-	1	-	-	-	-	-	-	-	-	+	-	112	2104
Synthia [70–72]	2016	2019	S	3C	-	-	1	-	-	D8	+	-	-	2	+	-	-	-	9400
Virtual KITTI [73,74]	2016	2020	S	2C	-	1	1	-	+	MD8	-	-	-	4	+	+	+	35	17k
MSSSD/MF [75]	2017	2017	R	1C/1T	-	-	-	-	-	DN	-	-	-	-	+	-	-	-	1569
RoadScene-Seg [76]	2018	2018	R	1C/1T	-	-	-	-	-	DN	-	-	-	-	-	-	-	-	221
AtUlm [77]	2019	2019	R	1G	4	-	-	-	-	-	-	-	-	-	+	-	-	-	1446
nuScenes [78]	2019	2020	R	6C	1	1	-	5	+	-	-	T	-	-	+	+	-	-	40k
SemanticKITTI [35]	2019	2021	R	-	1	-	-	-	-	-	-	-	-	-	+	-	-	22	43,552
ZJU [79]	2019	2019	R	2C/1FE/1T	-	1	-	-	-	DN	-	-	-	-	-	-	-	-	7400
A2D2 [80]	2020	2020	R	6C	5	-	-	-	+	-	-	-	-	-	+	+	-	-	41,280
ApolloScape [81]	2020	2020	R	6C	2	1	-	-	+	*	-	4R †	*	-	+	+	-	-	140k
DDAD [82]	2020	2020	R	6C	4	-	-	-	-	-	-	2R	-	-	-	-	-	-	16,600
KITTI 360 [83]	2021	2021	R	2C/2FE	1	1	-	-	+	-	-	-	-	-	+	+	-	-	78k
WoodScape [84]	2021	2021	R	4FE	1	-	-	-	+	-	-	10C	-	-	+	+	-	-	10k
EventScape [85]	2021	2021	S	1C/1E	-	-	1	-	+	-	-	4C	-	-	+	+	-	743(2 h)	-
SELMA [39]	2022	2022	S	8C	3	3	7	-	-	D8N	-	7C	9	+	+	+	-	-	31k x 27
Freiburg Forest [86]	2016	2016	R	2C/1MS	-	1	-	-	-	-	-	-	-	-	+	-	-	-	336
POLABOT [87]	2019	2019	R	2C/1F/1MS	-	1	-	-	-	-	-	-	-	-	+	-	-	-	175
SEM [88]	2021	2021	R	1C/1T	-	-	-	-	-	-	-	-	2	-	+	-	-	-	2458
SHV [89]	2021	2021	R	1C/1T	-	-	-	-	-	-	-	-	2	-	+	-	-	-	1573
MVSEC [90]	2018	2018	R	2G/2E	1	1	-	-	+	DN	-	IO	-	-	-	-	-	14(1h)	-
PST900 [91]	2019	2019	R	2C/1T	-	1	-	-	-	-	-	IO	-	-	+	-	-	-	4.3k
NYU-depth-v2 [92]	2012	2012	R	1C + 1D	-	-	1	-	-	-	-	-	-	-	+	-	-	-	1449 †
SUN-RGBD [93]	2015	2015	R	1C + 1D	-	-	1	-	-	-	-	-	-	-	+	-	-	-	10k †
2D-3D-S [94]	2017	2017	R	1C + 1D	-	-	1	-	-	-	-	-	-	-	+	-	-	-	270
ScanNet [95]	2017	2018	R	1C + 1D	-	-	1	-	-	-	-	-	-	-	+	-	-	-	1513
Taskonomy [77]	2018	2018	R	1C + 1D	-	-	1	-	-	-	-	-	-	-	-	-	-	-	4 m †

In short,

KITTI: The first extensive dataset to handle Multimodal information in self-driving cars is the KITTI vision benchmark, it was founded in 2012 with an emphasis on optical flow and semantic segmentation. Cityscapes: The cityscapes semantic segmentation dataset is frequently utilized for benchmarks in autonomous driving and depth estimate tasks. It includes five thousand high-resolution images of German cityscapes. Lost and Found: The 2016 dataset, which has over 2000 samples, emphasizes situations involving missing cargo using pixel-level segmentation of roadways and unnecessary things. 112 stereo video segments with 2,104 annotated frames are included. Synthia: Synthia is a 2016 multimodal synthetic dataset with 9400 samples that covers four seasons and day/night cycles and contains color, depth, and semantic data from multiple perspectives.

Virtual KITTI: is a synthetic Unity dataset that adds more labeled samples and more precision than the KITTI dataset, however, it has no LiDAR point cloud labels. MSSSD/MF: is a modest real-world dataset made up of 1.5k thermal cameras that serves as a standard for practical applications and provides multispectral information in low-visibility situations. RoadScene-Seg: This is a real-world dataset made up of 200 images of road scenes without labels. Architecture validation requires human qualitative review because the photos are label-free. AtUlm: is an actual dataset obtained from Ulm University from 2019 that includes 1446 carefully annotated samples that were obtained with four LiDARs and a grayscale camera. nuScenes: is a real-world dataset that serves as a baseline for structures utilizing such sensor modalities, providing RADAR data from five headlamp RADARs, one LiDAR, six top ring cameras, and an IMU. Semantic KITTI: is a point-wise labeling KITTI dataset expansion that has grew in favor as a benchmark for semantic segmentation due to the abundance of available samples. ZJU: provides a comprehensive image of the situation with 3400 labeled samples from a real-world dataset from 2019 that supports color, depth, light polarization modality, and fish-eye camera view. A2D2: is an actual dataset consisting of 41k samples from AUDI car manufacturers that emphasizes multimodal data, including cameras, LiDARs, weather variability, and IMU. ApolloScape: is a substantial real-world dataset that offers 150k annotated RGB photos together with depth data; nevertheless, its ability to be used in multimodal situations is limited by its static object maps. DDAD: is a monocular depth estimate dataset from the Toyota Research Institute that was captured in seven cities in the US and Japan; it only offers semantic segmentation labels for test sets and validation. KITTI 360: is a real-world dataset from 2020 that was collected using a synchronized sensor setup and contains 78K labeled samples from several modalities arranged in temporal sequences with little environmental variability. WoodScape: is a real-world dataset that extracts 2D information from more than ten cities spread over five different locales using fish-eye cameras. EventScape: is a synthetic dataset created in 2021 using the CARLA simulator. It contains information from 743 sequences spread across four cities, including color, depth, bounding boxes, event camera, semantic segmentation, and IMU. SELMA: is a synthetic dataset created in 2022 with an emphasis on semantic segmentation. It offers multimodal data in a range of environmental settings, giving researchers control over conditions and variability in the environment (Rizzoli, Barbato and Zanuttigh, 2022).

DOI:http://doi.org/10.24086/cocos2024/paper.1537

CONCLUSION

This paper presents an overview of semantic segmentation for autonomous driving due to its pivotal role in the advancement of autonomous vehicles, improving their capacity for accurate perception and interpretation of the surroundings. Autonomous vehicles can make defensible conclusions by breaking up an image into meaningful pieces and giving each region a semantic label based on a thorough grasp of the situation. A survey on multimodal object detection using different kinds of sensors for redundant and secure systems integration, a table comparison between different kinds of datasets that are available for this kind of segment, ranked according to the date of availability, an explanation graph listing sensing acquisition, approaches and methods categorized into Temporal model, context-aware model and fully conventional networks, some future research trends are listed for semantic segmentation for autonomous driving.

FUTURE DIRECTIONS IN RESEARCH:

Future research endeavors can utilize the promising methods presented in this section to enhance the performance of semantic picture segmentation (Papadeas et al., 2021):

- Domain adaptation: it is a transfer learning method that uses information from a different, related source domain to improve a model's performance on a target domain. It can improve autonomous driving's low latency and real-time semantic segmentation accuracy.

- Transfer learning: enhances domains with insufficient training data by transferring information from the source domain to the target domain. It offers better regularization and increased accuracy in semantic segmentation by lowering the amount of training data and time needed for real-time semantic segmentation.

- Self-supervised learning: is a subclass of unsupervised learning that learns representations from huge datasets without explicitly labeled data, avoiding human intervention in large-scale data annotation, especially in autonomous driving.

- Weakly supervised learning: is typically less expensive than fine-grained labels, and is typified by coarse-grained or erroneous labels. It performs better with classifier heatmaps and two-stream networks, and it achieves better accuracy results with the Cityscapes and CamVid datasets.

- Transformers: provide a worldwide approach to context modeling, outperforming convolutional techniques on ADE20K, by 4.6%, indicating their potential for future use in semantic segmentation.

REFERENCES

Feng, D. et al. (2021) 'Deep Multi-Modal Object Detection and Semantic Segmentation for Autonomous Driving: Datasets, Methods, and Challenges', IEEE Transactions on Intelligent Transportation Systems, 22(3), pp. 1341–1360. Available at: https://doi.org/10.1109/TITS.2020.2972974.

Lateef, F. and Ruichek, Y. (2019) 'Survey on semantic segmentation using deep learning techniques', Neurocomputing, 338. Available at: https://doi.org/10.1016/j.neucom.2019.02.003ï.

Muhammad, K. et al. (2022) 'Vision-Based Semantic Segmentation in Scene Understanding for Autonomous Driving: Recent Achievements, Challenges, and Outlooks', IEEE Transactions on Intelligent Transportation Systems, 23(12), pp. 22694–22715. Available at: https://doi.org/10.1109/TITS.2022.3207665.

Papadeas, I. et al. (2021) 'Real-time semantic image segmentation with deep learning for autonomous driving: A survey', Applied Sciences (Switzerland), 11(19). Available at: https://doi.org/10.3390/app11198802.

Rizzoli, G., Barbato, F. and Zanuttigh, P. (2022) 'Multimodal Semantic Segmentation in Autonomous Driving: A Review of Current Approaches and Future Perspectives', Technologies. MDPI. Available at: https://doi.org/10.3390/technologies10040090.

Siam, M. et al. (2017) 'Deep Semantic Segmentation for Automated Driving: Taxonomy, Roadmap and Challenges'. Available at: http://arxiv.org/abs/1707.02432.

Siam, M. et al. (no date) A Comparative Study of Real-time Semantic Segmentation for Autonomous Driving. Available at: https://github.com/MSiam/TFSegmentation.

Yurtsever, E. et al. (2020) 'A Survey of Autonomous Driving: Common Practices and Emerging Technologies', IEEE Access, 8, pp. 58443–58469. Available at: https://doi.org/10.1109/ACCESS.2020.2983149.

Zhang, J. et al. (2019) 'A Review of Deep Learning-Based Semantic Segmentation for Point Cloud', IEEE Access. Institute of Electrical and Electronics Engineers Inc., pp. 179118–179133. Available at: https://doi.org/10.1109/ACCESS.2019.2958671.

Long, J., Shelhamer, E. and Darrell, T. (no date) Fully Convolutional Networks for Semantic Segmentation, Available at: https://openaccess.thecvf.com/content_cvpr_2015/papers/Long_Fully_Convolutional_Networks_2015_CVPR_paper.pdf

Ronneberger, O., Fischer, P. and Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation. In Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18 (pp. 234-241). Springer International Publishing.

Zhang, H., Geiger, A. and Urtasun, R., 2013. Understanding high-level semantics by modeling traffic patterns. In Proceedings of the IEEE international conference on computer vision (pp. 3056-3063).

Testolina, P., Barbato, F., Michieli, U., Giordani, M., Zanuttigh, P. and Zorzi, M., 2023. SELMA: Semantic large-scale multimodal acquisitions in variable weather, daytime and viewpoints. IEEE Transactions on Intelligent Transportation Systems.

DOI:http://doi.org/10.24086/cocos2024/paper.1537