Thermal Image-Based Multi-Class Semantic Segmentation for Autonomous Vehicle Navigation in Restricted Environments

Nurul Fazri
Department of Electrical Engineering, Faculty of Engineering
Universitas Garut
Garut, Indonesia
24052320032@fteknik.uniga.ac.id

Helfy Susilawati
Department of Electrical Engineering, Faculty of Engineering
Universitas Garut
Garut, Indonesia
helfy.susilawati@uniga.ac.id

Mokh. Mirza Etnisa Haqiqi
Electrical Engineering Department, Faculty of Engineering
Universitas Indonesia
Depok, Indonesia
Mokhammad.mirza21@gmail.com

Arief Suryadi Satyawan
Research Center for Telecommunication
The National Research and Innovation Agency (BRIN)
Bandung, Indonesia
arie021@brin.go.id

Abstract—Technological advancements have propelled the development of environmentally friendly transportation, with autonomous vehicles (AVs) and thermal imaging playing pivotal roles in achieving sustainable urban mobility. This study explores the application of the SegNet deep learning architecture for multi-class semantic segmentation of thermal images in constrained environments. The methodology encompasses data acquisition using a thermal camera in urban settings, annotation of 3,001 thermal images across 10 object classes, and rigorous model training with a high-performance system. SegNet demonstrated robust learning capabilities, achieving a training accuracy of 96.7% and a final loss of 0.096 after 120 epochs. Testing results revealed strong performance for distinct objects like motorcycles (F1 score: 0.63) and poles (F1 score: 0.84), but challenges in segmenting complex patterns such as buildings (F1 score: 0.34) and trees (F1 score: 0.42). Visual analysis corroborated these findings, highlighting strengths in segmenting well-defined objects while addressing difficulties in handling variability and elongated structures. Despite these limitations, the study establishes SegNet's potential for thermal image segmentation in AV systems. This research contributes to the advancement of computer vision in autonomous navigation, fostering sustainable and green transportation solutions while emphasizing areas for further refinement to enhance performance in complex environments.

Keywords— Semantic Segmentation, SegNet, Autonomous Vehicles**,** Thermal Imaging

I. INTRODUCTION

Technological innovations play a critical role in enabling environmentally friendly or green transportation, with electric vehicles emerging as a key factor in achieving a green economy and sustainable development. Electric vehicles exemplify how technology can positively influence both the transportation industry and the natural environment [1]. Moreover, the development of electric vehicles can be advanced into autonomous vehicles capable of operating independently along predetermined routes and distances while prioritizing safety and comfort [2].

Autonomous vehicles represent a major innovation in the automotive industry, revolutionizing how humans drive. These vehicles can accelerate, navigate, detect their surroundings, avoid obstacles, and stop autonomously [3],[4]. Autonomous driving systems rely on the Prediction function to recognize and understand their environment, and the Perception function to avoid collisions and detect obstacles. These functions are supported by sensors such as LiDAR, cameras, radar, and GPS, with cameras playing a critical role in providing visual perception [5],[6].

Cameras use light sensors in a matrix format with specific dimensions and resolutions (Pixels) to capture images received through the lens. The resulting images must adapt to lighting conditions through ISO and brightness adjustments. However, in outdoor settings, challenges such as low light, sunlight glare, or fog can affect performance [7],[8]. To address these limitations, thermal cameras offer a solution as they can detect and measure infrared radiation emitted by objects, making them effective in conditions where traditional light sensors fail due to lighting or atmospheric challenges [9],[10]. The data captured represent infrared energy detected by thermal cameras. For thermal camera detection systems to be implemented in autonomous vehicles, reliable image processing and object recognition methods are required, particularly in constrained environments.

This study proposes an object detection system using thermal image-based segmentation techniques in constrained environments, employing semantic segmentation to deeply distinguish objects within images [11]. The system focuses on multi-class segmentation to detect 10 object classes, including road surfaces, skies, buildings, trees, and vehicles, leveraging the CNN method for classification and pattern recognition [12]. Previous studies, such as [13], utilized the ResNet 34 architecture with 40,216 thermal frames, while [14] implemented the ResNeXt 50 architecture with 34,030 thermal frames in the Universitas Nurtanio Bandung environment. This research takes a different approach by employing the SegNet architecture and a test dataset of 3,001 thermal frames in the constrained environment of the BRIN (National Research and Innovation Agency) office in Bandung. Based on other references, such as [15], which utilized the SegNet model for stereo camera image segmentation of obstacle objects with real-time (online) testing, an intriguing study [16] performed Road and Object Segmentation for Autonomous Vehicles using an RGB-D-NIR camera with 12 architectures, including SegNet. The key distinction of this research lies in the use of a thermal camera, with object annotation conducted for 10 object classes and implemented offline, resulting in a fundamentally different system.

The authors evaluated the CNN architecture using SegNet [17],[18] which has been proven effective for image segmentation, as evidenced by [19] for semantic segmentation efficiency, in [20] for high-accuracy railway segmentation, and [21] SegNet is effective and accurate in segmenting Fatty Liver Disease from ultrasound images, achieving an accuracy of 91%. This demonstrates its potential for multi-class segmentation in thermal images.

The implementation of the SegNet architecture for thermal image segmentation in constrained environments offers a novel approach to autonomous navigation. This study not only continues the exploration of thermal imaging for object detection, as mentioned in [22], but also extends its application to multi-class segmentation for a comprehensive understanding of environments. Constrained environments serve as controlled test areas to evaluate the effectiveness of thermal segmentation for various object classes, with the potential for broader applications in autonomous vehicle systems.

II. METHODOLOGY

A. Experimental Design

Figure 1 illustrates a detailed workflow diagram outlining the semantic segmentation process utilizing deep learning techniques. The approach encompasses several sequential stages, starting with data acquisition and concluding with model evaluation.

The first phase emphasizes data collection and preprocessing. Thermal video footage is recorded using a thermal camera positioned in a controlled environment at the BRIN (The National Research and Innovation Agency) Bandung Office. This setting primarily includes roads and the surrounding urban infrastructure. The recorded thermal videos are processed by extracting frames, resulting in a sequence of thermal images, each maintaining a resolution of 640x512 pixels.

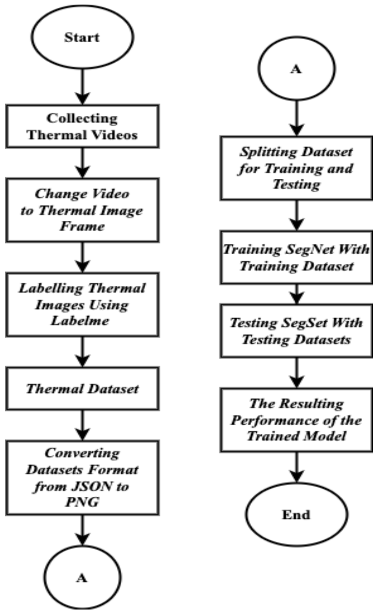
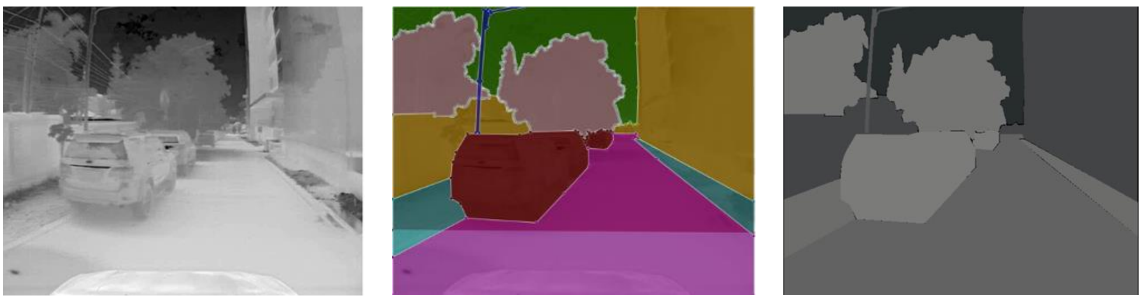


Figure 1. The Flow Diagram of The Experimental

The following annotation phase utilizes the LabelMe software to conduct precise object labeling across ten distinct categories: background, sky, building, tree, road, pavement, car, motorcycle, pedestrian, and pole. This thorough annotation procedure was applied to 3,001 thermal images, generating a comprehensive dataset for semantic segmentation tasks. The annotated data, initially stored in JSON format, is subsequently converted to PNG format to ensure compatibility with deep learning frameworks. During the dataset preparation phase, a strategic split is implemented, with 90% (2,700 images) allocated for model training and 10% (301 images) set aside for testing. This distribution guarantees an adequate amount of data for both model training and independent performance evaluation. Figure 2 shows an illustration of the dataset image that was used.


(a) (b) (c)
Figure 2. The Image for (a) Thermal Image, (b) Data Annotation, (c) PNG Format

Next, the training process employs the SegNet deep learning architecture. Model training is carried out on a high-performance computing system featuring an Nvidia

RTX 3060 GPU, running for 120 epochs to ensure adequate model convergence and optimal parameter adjustment. The workflow concludes with an extensive evaluation phase, where the trained SegNet model is tested using the reserved testing dataset to assess its performance in semantic segmentation. This methodical approach enables an objective evaluation of the SegNet architecture and confirms its effectiveness in thermal image segmentation tasks.

This methodology guarantees reproducibility while upholding scientific rigor throughout the experimental procedure, from data acquisition to model evaluation. The structured approach supports a detailed analysis of SegNet's capabilities in processing thermal imagery for autonomous navigation applications.

B. Parameters Output Models

The parameter output tables reveal distinctive architectural characteristics of SegNet (Table 1), showcasing its unique approach to semantic segmentation. The following examines the key features and architectural philosophy of SegNet in detail.

Table 1.The Parameter Output of The Architecture of SegNet

Layer	Layer (type)	Output Shap	Param
Input Stage	input_1 (InputLayer)	[(None, 416, 608, 3)]	0
	zero_padding2d (ZeroPadding 2D)	(None, 418, 610, 3)	0
First Layer	conv2d (Conv2D	(None, 416, 608, 64)	1792
	batch_normalization (BatchNormalization)	(None, 416, 608, 64)	256
	activation (Activation)	(None, 416, 608, 64)	0
	max_pooling2d (MaxPooling2D)	(None, 208, 304, 64)	0
	zero_padding2d_1 (ZeroPadding2D)	(None, 210, 306, 64	0
Second Layer	conv2d_1 (Conv2D)	(None, 208, 304, 128	73856
	batch_normalization_1 (Batc hNormalization)	(None, 208, 304, 128)	512
	activation_1 (Activation)	(None, 208, 304, 128)	0
	max_pooling2d_1 (MaxPooling 2D)	(None, 104, 152, 128)	0
	zero_padding2d_2 (ZeroPaddi ng2D)	(None, 106, 154, 128)	0
Third Layer	conv2d_2 (Conv2D)	(None, 104, 152, 256)	295168
	batch_normalization_2 (BatchNormalization)	(None, 104, 152, 256)	1024
	activation_2 (Activation)	(None, 104, 152, 256)	0
	max_pooling2d_2 (MaxPooling2D)	(None, 52, 76, 256)	0
	zero_padding2d_3 (ZeroPaddi ng2D)	(None, 54, 78, 256)	0
Fourth Layer	conv2d_3 (Conv2D)	(None, 52, 76, 256)	590080
	batch_normalization_3 (BatchNormalization)	(None, 52, 76, 256)	1024
	activation_3 (Activation)	(None, 52, 76, 256	0
	max_pooling2d_3 (MaxPooling2D)	(None, 26, 38, 256)	0
	zero_padding2d_4 (ZeroPadding2D)	(None, 28, 40, 256	0
Fifth Layer	conv2d_4 (Conv2D)	(None, 26, 38, 512)	1180160
	batch_normalization_4 (BatchNormalization)	(None, 26, 38, 512)	2048
	up_sampling2d (UpSampling2D)	(None, 52, 76, 512)	0
	zero_padding2d_5 (ZeroPadding2D)	(None, 54, 78, 512)	0
Sixth Layer	conv2d_5 (Conv2D	(None, 52, 76, 256)	1179904
	batch_normalization_5 (BatchNormalization)	(None, 52, 76, 256)	1024
	up_sampling2d_1 (UpSampling2D)	(None, 104, 152, 256)	0
	zero_padding2d_6 (ZeroPadding2D)	(None, 106, 154, 256)	0
Seventh Layer	conv2d_6 (Conv2D)	(None, 104, 152, 128	295040
	batch_normalization_6 (BatchNormalization)	(None, 104, 152, 128)	512
	up_sampling2d_2 (UpSampling2D)	(None, 208, 304, 128)	0
	zero_padding2d_7 (ZeroPadding2D)	(None, 210, 306, 128)	0
Eighth Layer	conv2d_7 (Conv2D)	(None, 208, 304, 64)	73792
	batch_normalization_7 (BatchNormalization)	(None, 208, 304, 64)	256
Ninth Layer	conv2d_8 (Conv2D)	(None, 208, 304, 10	5770
	reshape (Reshape)	(None, 63232, 10)	0
	activation_8 (Activation)	(None, 63232, 10	0

The SegNet architecture is an efficient encoder-decoder model designed for semantic segmentation tasks, offering a structured framework to process high-dimensional input tensors into pixel-wise class predictions. The network begins with an input tensor of size [416, 608,3], representing the height, width, and channels of the image. The first stage involves zero-padding, preserving the spatial dimensions while introducing no additional parameters. This step ensures consistent dimensions across subsequent layers.

In the encoder, each layer consists of convolutional operations, batch normalization, and max-pooling. The convolutional layers are characterized by kernels of size 3×3, with the number of parameters computed using the formula:

$$Params_{Conv2D} = (F_h . F_w . C_{in} + 1) . C_{out} \tag{1}$$

Where F_h *and* F_w are the kernel dimensions, C_{in} is the number of input channels, and C_{out} is the number of output channels. Following each convolution, batch normalization is applied, introducing $2 \ . \ C_{out}$ additional parameters to stabilize training. Dimensionality reduction is achieved via max-pooling, where spatial dimensions are halved according to the formula:

$$H_{out} = \frac{H_{in}}{P_s}, \quad W_{out} = \frac{W_{in}}{P_s}, \tag{2}$$

With $P_s = 2$ for a typical pooling size of 2×2. This step reduces the computational cost while preserving the most salient features of the input. The decoder mirrors the encoder in structure but substitutes max-pooling with upsampling layers to restore spatial resolution. Upsampling increases dimensions by a factor of two, governed by the formula:

$$H_{out} = H_{in} . U_s, \quad W_{out} = W_{in} . U_s \tag{3}$$

Where U_s is the upsampling factor, typically set to 2. The decoder's convolutional layers apply the same parameterization formula as the encoder, ensuring a symmetrical design.

In the final stage, the network reshapes the output tensor into a 2D matrix of size (N, K), where $N = H.W$ is the number of pixels in the image and K is the number of segmentation classes. This is followed by a softmax activation function, which assigns a probability distribution across all classes for each pixel:

$$Softmax(z_k) = \frac{e^{z_k}}{\sum_{j=1}^{K} e^{z_j}} \tag{4}$$

Where z_k is the logit for class k. This formulation allows the model to predict pixel-wise class probabilities for semantic segmentation tasks. The total number of parameters in the network is the summation of all parameters across convolutional and batch normalization layers:

$$Total\ Parameters = \sum_{i=1}^{n} Params_{Conv2D} + Params_{BatchNorm} \tag{5}$$

where n is the total number of convolutional layers.

III. Result and discussion

The analysis of the training accuracy and loss graph (Figure 3) from the SegNet architecture provides detailed insights into its learning process during the training phase using a dataset of 2,700 images covering 10 object classes.

The analysis of the training accuracy and loss graph from the SegNet architecture provides detailed insights into its learning process during the training phase, utilizing a dataset of thermal images for segmentation tasks across multiple object classes.

The training accuracy graph demonstrates a clear upward trend as the epochs progress. SegNet begins the training process with an initial accuracy of approximately 0.69 and shows steady improvement over the course of the training. The model achieves significant accuracy gains within the first 20 epochs, surpassing 90% by epoch 20 and continuing to improve gradually. By the end of training at epoch 120, the model reaches a final accuracy of approximately 0.967, indicating a high level of performance. The stabilization of the accuracy curve in later epochs suggests that SegNet successfully converges and achieves optimal learning without significant variability.

The training loss graph provides additional insight into SegNet's optimization process. Starting with an initial loss of about 0.927, the loss decreases sharply during the early epochs, reflecting rapid parameter adjustments and effective error minimization. By epoch 50, the loss drops below 0.2, and it continues to decline gradually, stabilizing at approximately 0.096 by epoch 120. This consistent reduction in loss demonstrates the model's ability to effectively optimize for the given segmentation task.

The combined analysis of the accuracy and loss trends underscores SegNet's reliable and stable learning characteristics. The simultaneous stabilization of accuracy and loss during the later epochs highlights proper convergence and indicates the absence of overfitting. The chosen training duration of 120 epochs proves to be sufficient for achieving robust results, enabling SegNet to handle the demands of thermal image segmentation with high precision and reliability.

The training results highlight SegNet's strong learning capabilities and consistent optimization process. The architecture demonstrates its suitability for thermal image segmentation tasks, handling the challenges effectively and producing accurate, stable, and convergent learning patterns throughout the training phase.

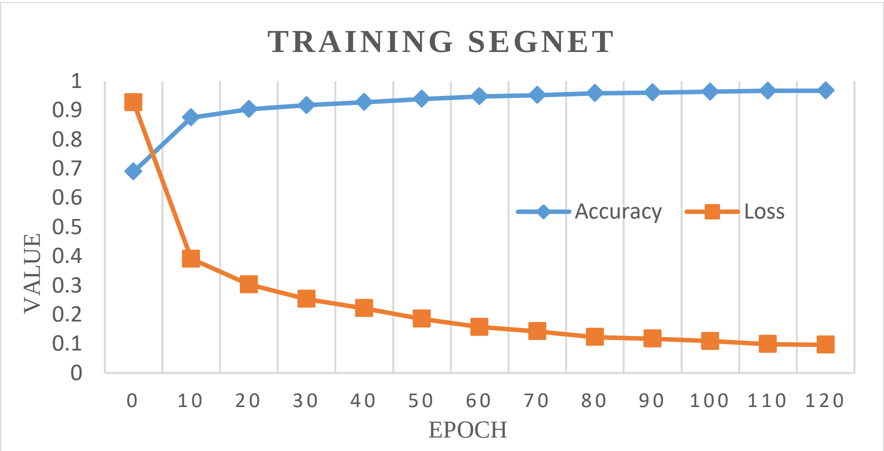


Figure 3. Training accuracy and loss given by SegNet after training process

The performance analysis of the SegNet model was conducted through testing on 301 thermal images containing 10 distinct object classes, following a training phase that utilized 2,700 thermal images. The evaluation metrics provided in Table 2 reveal the model's performance across multiple dimensions of assessment.

Tabel 2. Performance metrics given by SegNet

Class	**SegNet**				
	Accuracy	**Precision**	**Recall**	**F1 Score**	**AP**
Background	0.82	0.21	0.58	0.31	0.5
Sky	0.92	0.74	0.66	0.69	0.3
Building	0.69	0.37	0.32	0.34	0.35
Pole	0.95	0.77	0.93	0.84	0.18
Road	0.83	0.34	0.57	0.42	0.44
Pavement	0.76	0.53	0.58	0.55	0.79
Tree	0.75	0.69	0.28	0.4	0.89
Car	0.88	0.43	0.55	0.48	0.36
Pedestrian	0.93	0.72	0.72	0.72	0.37
Motorcycle	0.95	0.88	0.49	0.63	0.37
Average	0.85	0.57	0.57	0.54	0.46

SegNet demonstrates a varied performance across primary metrics in semantic segmentation tasks, as reflected in the provided evaluation results. On average, the model achieves 0.85 accuracy, 0.57 precision, 0.57 recall, 0.54 F1 score, and 0.46 average precision (AP). These metrics highlight SegNet's capability to handle segmentation tasks while also revealing limitations in specific areas.

A detailed analysis of class-specific performance shows that SegNet excels in detecting poles, pedestrians, and motorcycles, achieving accuracies of 0.95, 0.93, and 0.95, respectively. These results suggest that the model is effective in handling distinct and well-defined objects. However, the model faces challenges with classes such as buildings and trees, which record lower accuracies of 0.69 and 0.75, respectively. These difficulties are further reflected in low recall values for trees (0.28) and buildings (0.32), indicating the model's struggle to identify all relevant instances of these classes.

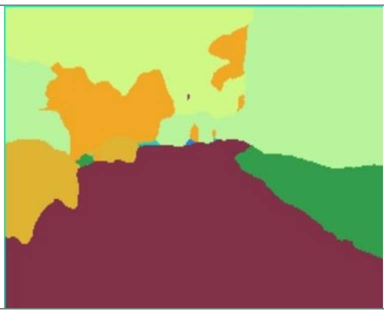
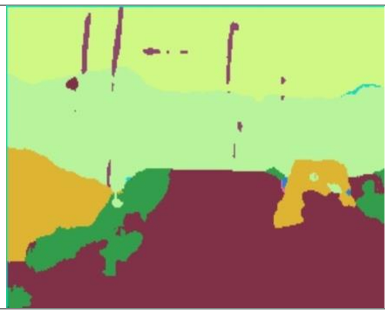
Precision scores point to significant challenges with false positives, particularly for the background (0.21) and building (0.37) classes, which indicate frequent misclassifications. Conversely, the model demonstrates stronger precision in detecting motorcycles (0.88) and poles (0.77), showing its robustness in these categories. The F1 score, which balances precision and recall, further illustrates these trends, with the pole class achieving the highest score (0.84) and the building class lagging behind (0.34).

The model's performance in the background class, with a low precision of 0.21 and an F1 score of 0.31, suggests difficulties in handling the variability of background patterns, likely caused by diverse environmental factors. Similarly, the relatively low F1 scores for roads (0.42) and cars (0.48) highlight limitations in detecting safety-critical classes, which could impact the model's applicability in domains like autonomous driving or urban surveillance.

In contrast, SegNet shows strong potential in pedestrian detection (F1 score of 0.72) and motorcycle detection (F1 score of 0.63), making it particularly valuable for applications requiring accurate identification of dynamic objects. However, the low average precision (AP) scores for certain classes, such as poles (0.18) and background (0.50), indicate challenges in maintaining consistent detection performance across diverse scenarios.

SegNet demonstrates robust performance for specific object classes but faces notable challenges with false positives, low recall, and certain variable patterns. These results suggest the need for targeted architectural refinements and training strategies to improve the model's reliability and effectiveness in complex segmentation tasks.

Tabel 3. Visual outputs given by SegNet

	SegNet	
AP	Maximum: 0.89	Minimum: 0.18
Class	Tree	Pole
Segmented Image Prediction		

The visual outputs from SegNet (Table 3) provide key insights into its segmentation performance when combined with its quantitative metrics. A detailed review highlights how SegNet handles different object classes and boundary conditions.

For the "Tree" class, which achieves the highest AP score of 0.89, the visual outputs display well-defined boundaries and consistent region segmentation. These results align with the model's training metrics, particularly its high precision (0.69) for trees, indicating that SegNet effectively identifies and separates tree regions from the surrounding background.

In contrast, the "Pole" class, with the lowest AP score of 0.18, demonstrates significant challenges. The visual predictions show fragmented and incomplete pole structures, reflecting difficulties in detecting thin and elongated objects. This is consistent with the quantitative metrics, where poles exhibit high recall (0.93) but low precision, suggesting the model over-identifies similar structures in the background, leading to false positives.

The visual outputs highlight SegNet's strengths in segmenting clear and well-defined objects, like trees, but also its limitations with complex or narrow structures, like poles. These challenges emphasize the need for improved feature extraction and post-processing techniques to enhance segmentation accuracy across diverse object classes.

IV. CONCLUSION

This study demonstrates the potential of the SegNet architecture for multi-class semantic segmentation of thermal images in constrained environments, achieving a training accuracy of 96.7% and a stabilized loss of 0.096. While SegNet excels in segmenting well-defined objects such as poles, motorcycles, and pedestrians, challenges persist with complex patterns like buildings and trees. These findings highlight the need for enhanced feature extraction techniques and architectural improvements to address class-specific variability. By advancing the application of thermal imaging in autonomous navigation, this research provides a strong foundation for future developments in sustainable urban mobility and green technology.

ACKNOWLEDGEMENT

This research is supported by Faculty of Engineering, Universitas Garut which has provided support and a place for carrying out the research and supported by funding from the Indonesia Endowment Fund for Education (Lembaga Pengelola Dana Pendidikan) through the National Research and Innovation Agency's Research and Innovation for Advanced Indonesia (RIIM) program (BRIN).

REFERENCES

[1] R. D. Saniyyah, U. Islam, N. Sunan, and A. Surabaya, "Peran Inovasi Teknologi Dalam Green Transportasi : Mewujudkan Green Economy Dan Pembangunan," 2024.

[2] M. Rauf, L. Kumar, S. A. Zulkifli, and A. Jamil, "Aspects of artificial intelligence in future electric vehicle technology for sustainable environmental impact," *Environmental Challenges*, vol. 14, Jan. 2024, doi: 10.1016/j.envc.2024.100854.

[3] A. Tijani, "Obstacle Avoidance Path Design for Autonomous Vehicles – A Review," vol. 3, no. 5, pp. 64–81, 2021.

[4] D. Garikapati, "Autonomous Vehicles : Evolution of Artificial Intelligence and the Current Industry Landscape," 2024.

[5] S. More, A. Singh, P. Jana, and B. Pal, "Perception and Planning in Autonomous Car," no. December, 2020, doi: 10.13140/RG.2.2.31312.38401.

[6] A. B. Pratama, R. Effendi, A. Kadir, and A. Hady, "Deteksi Ruang Kosong pada Jalan Menggunakan Semantic Segmentation pada Mobil Otonom," vol. 11, no. 1, 2022.

[7] A. Silwal, T. Parhar, F. Yandun, and G. Kantor, "A Robust Illumination-Invariant Camera System for Agricultural Applications," Jan. 2021, doi: 10.1109/IROS51168.2021.9636542.

[8] T. Sefer and R. Ayaz, "Performance investigation of different headlights used in vehicles under foggy conditions," *Sci Rep*, pp. 1–12, 2023, doi: 10.1038/s41598-023-31883-3.

[9] J. Manuel *et al.*, "Analysis of Thermal Imaging Performance under Extreme Foggy Conditions : Applications to Autonomous Driving," 2022.

[10] A. N. Wilson, K. A. Gupta, B. H. Koduru, A. Kumar, A. Jha, and L. R. Cenkeramaddi, "Recent Advances in Thermal Imaging and its Applications Using Machine Learning: A Review," *IEEE Sensors Journal*, vol. 23, no. 4, pp. 3395–3407, 2023, doi: 10.1109/JSEN.2023.3234335.

[11] Y. Guo and B. Yang, "A Survey of Semantic Segmentation Methods in Traffic Scenarios," in *2022 International Conference on Machine Learning, Cloud Computing and Intelligent Mining (MLCCIM)*, 2022, pp. 452–457. doi: 10.1109/MLCCIM55934.2022.00083.

[12] X. Zhao, L. Wang, Y. Zhang, X. Han, and M. Deveci, *A review of convolutional neural networks in computer vision*, vol. 57, no. 4. Springer Netherlands, 2024. doi: 10.1007/s10462-024-10721-6.

[13] M. Laksono, A. Satyawan, and S. Siswanti, “Segmentasi Objek Berbasis Kamera Termal Menggunakan Deep Learning (Pre-Trained Resnet 34)Thermal Image-Based Object Segmentation Using Deep Learning (Pre-Trained Resnet 34),” *Prosiding Seminar Nasional Sains Teknologi dan Inovasi Indonesia (SENASTINDO)*, vol. 4, pp. 333–343, 2022, doi: 10.54706/senastindo.v4.2022.210.

[14] R. Fauzan, A. Satyawan, S. Siswanti, and H. Puspita, “Segmentasi Objek Berbasis Gambar Termal Menggunakan Deep Learning (Pre-Trained Resnext 50),” vol. 4, no. September, pp. 308–319, 2022.

[15] U. Ulusoy, O. Eren, and A. Demirhan, “Development of an obstacle avoiding autonomous vehicle by using stereo depth estimation and artificial intelligence based semantic segmentation,” *Engineering Applications of Artificial Intelligence*, vol. 126, p. 106808, 2023, doi: https://doi.org/10.1016/j.engappai.2023.106808.

[16] A. Helnawan, M. Attamimi, and A. N. Irfansyah, “Sistem Segmentasi Jalan dan Objek untuk Kendaraan Otonom Menggunakan Kamera RGB-D,” *Jurnal Teknik ITS*, vol. 12, no. 1, 2023, doi: 10.12962/j23373539.v12i1.110848.

[17] C. Zhang, W. Lu, J. Wu, C. Ni, and H. Wang, “SegNet Network Architecture for Deep Learning Image Segmentation and Its Integrated Applications and Prospects,” vol. 9, no. 2, 2024.

[18] E. Chamseddine, L. Tlig, M. Sayadi, and M. Bouchouicha, “SegNet Architecture for Dermoscopic Image Segmentation,” in *2022 IEEE Information Technologies & Smart Industrial Systems (ITSIS)*, 2022, pp. 1–6. doi: 10.1109/ITSIS56166.2022.10118404.

[19] M. Yumuş, M. Apaydın, A. Değirmenci, H. Kaplanoğlu, S. Kesikburun, and Ö. Karal, “Deep Convolutional Neural Networks Using SegNet for Automatic Spinal Canal Segmentation in Axial MRI,” in *2023 Innovations in Intelligent Systems and Applications Conference (ASYU)*, 2023, pp. 1–6. doi: 10.1109/ASYU58738.2023.10296627.

[20] S. Khrueakhrai and J. Srinonchat, “Railway Track Detection Based on SegNet Deep Learning,” in *TENCON 2023 - 2023 IEEE Region 10 Conference (TENCON)*, 2023, pp. 409–413. doi: 10.1109/TENCON58879.2023.10322378.

[21] G. Paul and G. Ramkumar, “An Automated Classification and Segmentation of Fatty Liver Disease Using SegNet Model on Ultrasound Images,” in *2023 International Conference on Emerging Research in Computational Science (ICERCS)*, 2023, pp. 1–6. doi: 10.1109/ICERCS57948.2023.10434261.

[22] K. R. Akshatha, A. K. Karunakar, S. B. Shenoy, and A. K. Pai, “Human Detection in Aerial Thermal Images Using Faster R-CNN and SSD Algorithms,” pp. 1–15, 2022.