ACTA IMEKO
ISSN: 2221-870X
September 2025, Volume 14, Number 3, 1 - 8

Impact of data augmentation on labelling confidence in deep learning terrain traversability analysis for unmanned ground vehicles

Sebastiano Chiodini[1,2], Giulio Polato[1,3], Andrea Valmorbida[1,2], Marco Pertile[1,2], Giada Giorgi[4], Claudio Narduzzi[4], Enrico C. Lorenzini[1,2]

[1] *Department of Industrial Engineering (DII), University of Padova, Padova, Italy*
[2] *CISAS "Giuseppe Colombo", University of Padova, Padova, Italy*
[3] *Department of Physics, University of Trento, Trento, Italy*
[4] *Department of Information Engineering (DEI), University of Padova, Padova, Italy*

ABSTRACT
This study investigates the impact of data augmentation techniques on the accuracy and prediction probability of deep learning-based terrain classification systems for Unmanned Ground Vehicles (UGVs) in unstructured environments. The challenge of limited datasets in such environments is addressed through the implementation and evaluation of various data augmentation methods, to enhance the accuracy and reliability of pixel-level terrain measurements. The methodology is based on the DeepLabv3+ neural network architecture for supervised learning, trained on a custom dataset collected from an outdoor environment. A systematic assessment of multiple augmentation strategies is conducted, including geometric transformations (cropping and mirroring), colour space modifications (HSV transformations), and noise injection (Gaussian noise addition). The performance of these techniques is quantified using standard metrics, such as classification accuracy and Intersection over Union (IoU), alongside an analysis of pixel-wise classification prediction probability. Results indicate that, while traditional metrics show modest improvements, the application of data augmentation significantly enhances the model's prediction probability in its measurements, particularly for critical terrain features, such as traversable paths. A detailed analysis of the prediction probability distribution is presented, showing a significant improvement in the model's confidence for correctly classified pixels. Specifically, when augmentation strategies are applied, the percentage of traversable terrain pixels classified with high confidence (> 99.7 % probability) significantly increased from 75 % to 85 %.

Section: RESEARCH PAPER

Keywords: data augmentation; deep learning; Unmanned Ground Vehicles (UGVs); autonomous navigation

Citation: S. Chiodini, G. Polato, A. Valmorbida, M. Pertile, G. Giorgi, C. Narduzzi, E. C. Lorenzini, Impact of data augmentation on labelling confidence in deep learning terrain traversability analysis for unmanned ground vehicles, Acta IMEKO, vol. 14 (2025) no. 3, pp. 1-8. DOI: 10.21014/actaimeko.v14i3.2067

Section Editor: Andrea Scorza, University Roma Tre, Italy

Received February 2, 2025; **In final form** August 30, 2025; **Published** September 2025

Copyright: This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Corresponding author: Sebastiano Chiodini, e-mail: sebastiano.chiodini@unipd.it

1. INTRODUCTION

Accurate terrain traversability analysis and robust obstacle detection remain critical challenges in the development of autonomous navigation systems, despite the significant advancements in artificial intelligence for self-driving technologies. Accidents involving autonomous vehicles have highlighted the persistent need for enhanced safety measures and more reliable obstacle detection algorithms [1]. These challenges extend beyond on-road vehicles, covering a broader range of autonomous systems, such as Unmanned Ground Vehicles (UGVs) for applications in smart agriculture, planetary exploration, and disaster response operations [2].

The state of the art in autonomous navigation relies on the availability of an occupancy grid map, where cells are divided into free and non-traversable areas, such as those occupied by obstacles. This approach is probabilistic in nature, as each cell is characterised by the probability of being occupied or not. Cells are typically considered "occupied" or "free", with probabilities ranging from 0 to 1 [3]. As a robot or a vehicle moves through the environment, the map is continuously updated based on sensor data, such as Light Detection and Ranging (LiDAR)

systems or stereo vision. LiDAR is a remote sensing technology that uses a laser pulse and measures the time it takes for the pulse to reflect off a surface and return to the sensor, a principle known as time-of-flight. This allows LiDAR systems to create a detailed 3D point cloud representation of the environment. Stereo vision, on the other hand, uses two or more cameras to capture overlapping images, which can be further processed to extract depth information and generate a 3D map. Map updates are carried out using the Bayesian probability theory, integrating new sensor readings to refine the occupancy probability of each cell [4]. Global and local path planners then use these occupancy grid maps for UGV trajectory determination [5].

By processing point clouds obtained from LiDAR or stereo vision, it is possible to identify obstacles obstructing the path. However, point cloud analysis alone is not sufficient for the traversability analysis of terrains where the vehicle might lose traction, such as ice, gravel, or slippery areas. For this reason, a detailed terrain analysis through semantic segmentation, where each pixel is labelled with a class, is necessary. Semantic segmentation thus becomes the core of a reliable autonomous navigation system [6].

Various methods have been proposed to aid in the safe navigation of UGVs. [7] proposed a vision-based method for classifying terrain, reaching 99 % accuracy by utilising innovative image features and classifiers, while [8] crafted visual classifiers for terrains, attaining 70–93 % accuracy through the use of simplified texture descriptors. These techniques are based on image classification, which focuses on assigning a single category label to the entire image, such as safe terrain or rock hazard. However, they lack detailed information about hazard boundaries. Instead, image-semantic segmentation delineates every pixel in the image with a specific label, precisely outlining hazards such as rocks and cliffs [9].

Deep learning architectures, such as Convolutional Neural Networks (CNN), have shown impressive performance in implementing this fine-grain image segmentation for terrain analysis[10]. These architectures yield more detailed information; however, they necessitate greater computational resources and extensive datasets to achieve high degrees of accuracy. Nevertheless, in sectors such as planetary exploration, agricultural robotics, and disaster response operations, the availability of data is frequently constrained. The requirement for resilient vision systems, which must work under diverse lighting conditions, worsens this issue. Data augmentation is identified as an efficacious strategy to address this problem, as outlined in [11] and [12]. Recent studies highlight the importance of reliable classification systems, not only for vehicle traversability analysis but also across various critical domains. [13] addresses the challenge of limited medical X-ray data for respiratory disease detection using CNNs, proposing a new deep learning approach. [14] explores neural network-based data augmentation to improve the classification of underrepresented defect classes in industrial quality assurance. In a related field, [15] proposes an improved RGB-D Mask R-CNN method for cabin pose measurement in large cabin assembly processes. These studies pose similar challenges to the problem of terrain traversability semantic segmentation. [16] introduces an innovative stereo-vision-based semantic segmentation framework for autonomous navigation, utilising a stereo camera setup similar to our work. However, while this study focuses on maximising the use of stereo image data through novel label generation and information fusion techniques, our research emphasizes a distinct aspect—analysing and improving the predicted confidence of the classification.

Building on this context, in the works [17], [18], and [19], we have developed a semantic segmentation method to identify various types of terrain and potential hazards in images captured by UGVs. Our approach involves assessing a vehicle or robot's ability to navigate a given terrain, based on its geometric characteristics and visual data. The proposed architecture combines: stereo visual SLAM (Simultaneous Localization and Mapping) to reconstruct the vehicle's trajectory [20], supervised learning with DeepLabv3+ to achieve fine-grain pixel-level classification [21], and Occupancy Grid Mapping [18]. However, it is worth noting that our previous works [17], [18], and [19], were primarily focused on the mapping architecture. In contrast, this current work shifts the focus on the impact of data augmentation techniques on image segmentation performance.

In real-world scenarios, classification networks must not only be accurate, but they should also indicate when they are likely to be incorrect. The ability to quantify the prediction probability is crucial, as it allows the system to make informed decisions based on the reliability of the outputs. Consider, for instance, a planetary rover traversing gravelly terrain. The rover uses a neural network to classify the surface ahead and determine if it is safe to proceed. If the network cannot confidently categorise an area as traversable, the rover should immediately replan its trajectory. It might choose to entirely avoid the uncertain area, opting instead for a path where the terrain classification is more reliable. This challenge of ensuring well-calibrated prediction confidence is further highlighted in [22], which examines the problem of confidence calibration in modern neural networks.

The literature presents several studies which evaluate the impact of data augmentation techniques on semantic segmentation methods, primarily focusing on traditional metrics such as Accuracy and Intersection over Union, as in the works of [11] and [12]. Additionally, there is a body of research dedicated to developing novel approaches for quantifying classification uncertainty, such as in [23] and [24]. In contrast to most existing studies, our research aims to elucidate the influence of data augmentation techniques on the uncertainty associated with classification outcomes. Specifically, the investigation concentrates on the effects of four key data augmentation methods: cropping, mirroring, HSV (Hue, Saturation, Value) colour space changes, and the addition of Gaussian noise.

The article is structured as follows: Section 2 presents the method, Section 3 covers the tests performed and their results, while Section 4 contains the conclusions of this work.

2. METHODS

To classify terrain, a Convolutional Neural Network (CNN) approach was implemented using the DeepLabv3+ architecture. DeepLabv3+ is a state-of-the-art architecture for semantic image segmentation, introduced by Liang-Chieh Chen et al. in [21]. It builds upon previous DeepLab versions, including an encoder-decoder module to refine segmentation results, particularly along object boundaries. The architecture uses atrous convolutions to increase the receptive field of neurons, without increasing the number of parameters, allowing the capture of large-scale contextual information. DeepLabv3+ is widely used in various computer vision tasks, including autonomous driving, medical diagnostics, and satellite image analysis, due to its ability to accurately segment images with fine details. This architecture was

Table 1. Data augmentation techniques employed in the generation of the training dataset. This table outlines the various data augmentation methods considered and applied to enhance the diversity and size of the original dataset.

	Cropping	**Cropping + Mirroring + HSV**	**Cropping + Mirroring + HSV + GN σ = 0.01**	**Cropping + Mirroring + HSV + GN σ = 0.1**	**Cropping + Mirroring + HSV + GN σ = 0.01 + GN σ = 0.1**
Cropping	✓	✓	✓	✓	✓
Mirroring		✓	✓	✓	✓
HSV Change		✓	✓	✓	✓
Gaussian Noise σ = 0.01			✓		✓
Gaussian Noise σ = 0.1				✓	✓
Number of Images	1014	4056	5070	5070	6084

chosen for its proven effectiveness, efficiency, and adaptability to different datasets and segmentation tasks [6].

This network was trained to carry out pixel-wise segmentation of input images, identifying traversable terrain areas. The network training followed a supervised learning process, using a dataset of representative images collected in a park environment. The training process was structured into four main phases: data labelling, data augmentation, network training, and network testing. For the sake of this study, the ResNet18 encoder was chosen from within the DeepLabv3+ framework to optimise training times.

A key aspect of the methodology was the implementation of various data augmentation techniques to enhance the diversity and size of the training dataset. These techniques included geometric transformations (cropping and mirroring), HSV (Hue, Saturation, Value) colour space changes, and the addition of Gaussian noise with different standard deviations. The data augmentation process was designed to improve the CNN's ability to generalise across various environmental conditions and imaging variations, with the goal of increasing the model's performance and reliability in terrain classification within complex, unstructured environments.

2.1. Labelling

The system relies on manually labelled data for five terrain types: paved (critical for navigation), grassland (important for specific terrains), other vegetation, cars, and unidentified (ensures complete labelling). 169 RGB images with a resolution of 376 × 672 pixels have been manually labelled to train the neural network and represent the original dataset. These images were randomly chosen from a sequence captured within the test area using a ZED stereo camera. It is worth noting that, while the ZED camera captures stereoscopic imagery, only one of the two stereo images (specifically, the left image) was used for this terrain classification task. The resolution of 376 × 672 pixels refers to these single, left-eye images used in the dataset. Due to the limited dataset size, data augmentation techniques are used to improve training data.

2.2. Data augmentation

As mentioned earlier, to increase the number of training images, the following augmentation techniques have been used to train the CNN: cropping, mirroring, HSV (Hue, Saturation, Value) change, Gaussian noise addition, and some combinations of these transformations.

The cropping technique involves the extraction of sections of an original image to produce new, smaller images for training purposes. This approach assists in managing variable input sizes and can address issues related to class imbalance in semantic segmentation tasks [25]. It was determined to partition the image into six squares in a deterministic manner, as shown in Figure 1.

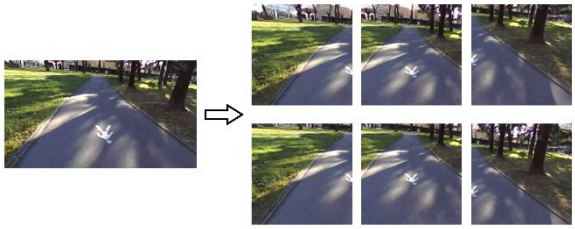

Figure 1. Example of data augmentation through cropping. The cropping technique involves cutting portions of an input image, while preserving its key features.

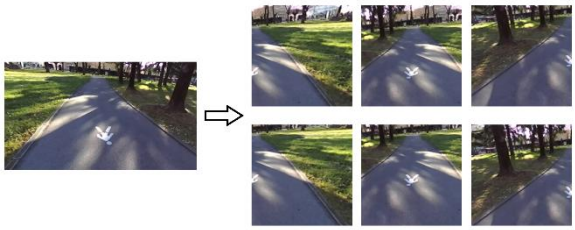

Figure 2. Example of data augmentation through image mirroring. The mirroring technique creates new training samples by flipping the original input image horizontally or vertically.

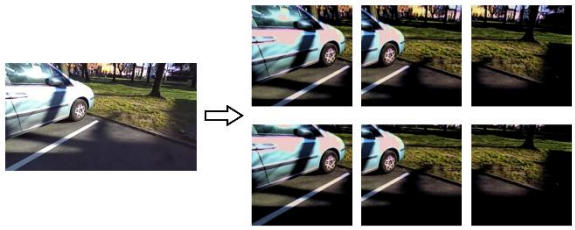

Figure 3. Example of data augmentation through HSV (Hue, Saturation, Value) colour space transformation. The HSV colour space transformation technique modifies the colour attributes of the input image, such as hue, saturation, and brightness, while preserving the overall structure.

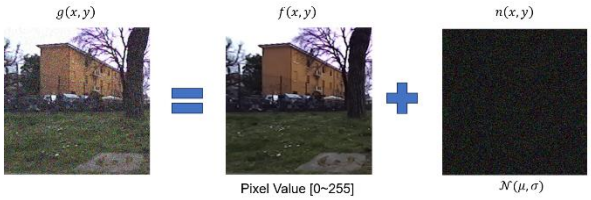

Figure 4. Example of data augmentation through Gaussian noise addition. The Gaussian noise technique involves adding random noise to an input image, simulating imperfections or distortions, while preserving its overall structure.

After the application of cropping, we obtained 1014 RGB images, each with dimensions of 224 × 224 pixels. This specific dimension of 224 × 224 pixels was chosen as it is the required input size for the first layer of the DeepLabv3+ architecture.

The mirroring technique is a simple yet effective data augmentation technique that involves the creation of a mirror image from the original one, typically along the horizontal or vertical axes. Mirroring improves model generalisation by exposing the network to different orientations of objects and scenes [26]. After applying mirroring on the original image, cropping is applied again, as the network receives input images of a fixed dimension. In this way, we again obtain 1,014 RGB images with a dimension of 224 × 224 pixels. Figure 2 shows an example of data augmentation through image mirroring. By applying cropping without mirroring and cropping with mirroring, the total number of images available for training is 2,028.

The HSV change technique is applied to enhance the colour contrast of the image by adjusting the intensity ratio of hue, saturation, and value channel. By simulating different lighting conditions and colour variations, HSV augmentation helps models learn features that are invariant to colour changes [27]. Notably, the HSV change technique adjusts the colour of the RGB image by adding a random amount of brightness, selected from a uniform distribution in the range [-0.3, 0.3], a random contrast scaling factor chosen from the uniform distribution [0.6, 1.4], and adding a random amount of saturation picked from the uniform distribution [-0.2, 0.2]. HSV change technique is applied on the original data set in this case as well, which is then cropped to obtain again 1,014 RGB images with a dimension of 224 × 224 pixels. Figure 3 shows an example of data augmentation through HSV change. By applying cropping without mirroring, cropping with mirroring, cropping with HSV change, and cropping with mirroring and HSV change, the total number of images available for training is 4,056.

The Gaussian noise (GN) addition technique is a data augmentation technique that introduces random noise to the training images. With the aim of increasing the variability of the images used for training and simulating sensor noise, we chose to apply this technique under two different conditions, to investigate the impact of varying noise levels on segmentation accuracy. Notably, we used two different standard deviation levels: $\sigma = 0.01$ and $\sigma = 0.1$. This approach allows us to evaluate how the model's performance changes with lower and higher levels of added noise. These two levels of Gaussian noise were applied separately, resulting in two training datasets of 5,070 images each. Subsequently, we applied both levels of Gaussian noise together, resulting in a combined training dataset of 6,084 images. Figure 4 shows an example of data augmentation through Gaussian noise addition.

Incorporating Gaussian noise during training helps the model learn more robust and generalised features, making it less sensitive to minor variations in the input data. Additionally, Gaussian noise augmentation can help prevent overfitting by forcing the model to learn more general patterns, rather than relying on specific image details [28]. In this case as well, it is necessary to apply cropping after Gaussian noise addition to the initial image. By applying all the data augmentation techniques discussed so far, the total number of images available for training is 6,084. Table 1 outlines the various data augmentation methods considered and applied to enhance the diversity and size of the original dataset.

2.3. Network training and testing

The training process leverages the SGDM (Stochastic Gradient Descent with Momentum) algorithm for optimisation. SGDM is a widely used optimiser in machine learning tasks; it is an extension of the standard stochastic gradient descent algorithm that incorporates a momentum term.

This momentum helps accelerate the optimisation process by accumulating a velocity vector in directions of persistent reduction in the objective function across iterations [29]. The available dataset is partitioned into three distinct groups – Training (70 %), Validation (15 %), and Testing (15 %) – following the common deep learning practice for model development [30].

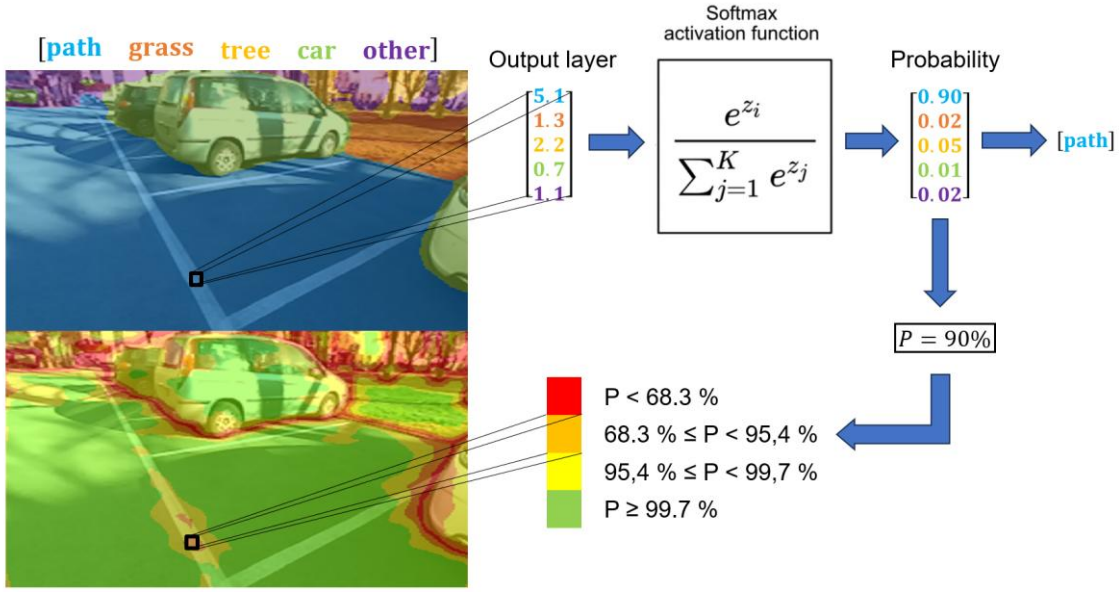

Figure 5. Pixel-wise terrain classification using semantic segmentation. This figure illustrates the process of assigning labels to individual pixels in the input image. The neural network's final layer outputs are passed through a Softmax activation function, which converts the raw output values into a probability prediction across the predefined terrain classes. Each pixel is then labelled according to the class with the highest probability. The color-coded overlay on the image represents different terrain types, demonstrating how the model classifies each pixel based on the Softmax output.

Table 2. Accuracy, Intersection over Union (IoU), and F1 score metrics for the class of interest after training, using various combinations of data augmentation techniques.

	Cropping			Cropping + Mirroring + HSV			Cropping + Mirroring + HSV + GN σ=0.01			Cropping + Mirroring + HSV + GN σ=0.1			Cropping + Mirroring + HSV + GN σ=0.01 + GN σ=0.1		
	Acc.	IoU	F1	Acc.	IoU	F1	Acc.	IoU	F1	Acc.	IoU	F1	Acc.	IoU	F1
Path	0.98	0.98	0.99	0.98	0.97	0.98	0.98	0.97	0.98	0.98	0.96	0.98	0.97	0.96	0.98
Grassland	0.97	0.95	0.98	0.97	0.95	0.98	0.96	0.94	0.97	0.97	0.95	0.97	0.97	0.95	0.97
Other Vegetation	0.91	0.70	0.82	0.88	0.67	0.80	0.92	0.65	0.78	0.88	0.65	0.79	0.94	0.64	0.78
Car	0.96	0.88	0.93	0.96	0.87	0.93	0.97	0.87	0.93	0.97	0.82	0.90	0.96	0.88	0.94
Unidentified	0.92	0.89	0.94	0.91	0.87	0.93	0.90	0.86	0.93	0.89	0.85	0.92	0.87	0.84	0.91

In our study, we evaluated the classic metrics of accuracy, Intersection over Union (IoU), and F1-score as a function of the applied data augmentation techniques. This approach allowed us to assess the impact of various augmentation methods on the model's performance across these standard metrics [28]. Accuracy measures the overall correctness of predictions, and it is calculated as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}, \quad (1)$$

where *TP*, *TN*, *FP*, and *FN* are respectively the true positives, the true negatives, the false positives, and the false negatives. IoU is commonly used in object detection and segmentation, measuring the overlap between predicted and ground truth bounding boxes or masks, and it is calculated as follows:

$$IoU = \frac{TP}{TP + FP + FN}. \quad (2)$$

The F1-score provides a balanced measure of a model's performance, especially useful for imbalanced datasets, and it is calculated as follows:

$$F1 = \frac{2TP}{2TP + FP + FN}. \quad (3)$$

Furthermore, we examined the confidence with which individual pixels are labelled. This analysis was made possible by leveraging the SoftMax layer available in the DeepLabv3+ architecture. The SoftMax activation function, added as the final layer to the neural network, converts the raw output into prediction probabilities [20]. This allowed us to not only classify pixels, but also to quantify the model's certainty in its pixel-wise predictions.

Figure 5 shows an example of pixel-wise terrain classification using semantic segmentation, along with the associated confidence estimate for each classification. The neural network output for each pixel consists of an array of scalars, indicating the pixel's similarity to each considered class. This array is then normalised through the SoftMax function, resulting in a pseudo-probability distribution of class membership for each pixel. The SoftMax activation layer converts a vector of unnormalised predictions $\boldsymbol{z}$ into a vector $\sigma(\boldsymbol{z})_\mathrm{i}$ of probabilities that sum to 1. The SoftMax function is defined as:

$$\sigma(\boldsymbol{z})_\mathrm{i} = \frac{\mathrm{e}^{z_\mathrm{i}}}{\sum_{\mathrm{k}=1}^{K} \mathrm{e}^{z_\mathrm{k}}}, \quad (4)$$

where $\sigma(\boldsymbol{z})_\mathrm{i}$ is the SoftMax output for class i, z_i is the input for class i, and the denominator sums over all classes k. This normalisation ensures that the outputs are between 0 and 1, and sum to 1, analogous to a probability distribution. The exponential nature of SoftMax accentuates larger input values, while suppressing smaller ones, making the predicted class probabilities more distinct. In the context of terrain classification, the SoftMax layer enables the network to assign probabilities to different terrain types for each pixel, facilitating informed decision-making in autonomous navigation systems [32]. As the predicted probability of the class to which a single pixel belongs, the maximum value of SoftMax output $\sigma(\boldsymbol{z})_\mathrm{i}$ over all classes k was taken. This approach allows for a definitive classification of each pixel based on the highest probability output by the SoftMax function.

2.4. Experimental setup

The images in the dataset were collected using a ZED stereo camera, which is widely employed in the field of navigation for UGVs. The camera was set to a resolution of 376 × 672 pixels, balancing image quality with computational efficiency. This camera choice and resolution setting reflect common practices in autonomous navigation research and applications [33]. The trained network underwent evaluation on a sequence of 200 images that were captured in the same location as the training images, albeit different from those used during the training phase.

3. TEST AND RESULTS

To evaluate the effectiveness of various data augmentation techniques, we conducted a series of experiments using different combinations of these methods. Table 2 presents the Accuracy, IoU, and F1 score metrics for each class of interest after training, under different augmentation scenarios. The augmentation techniques considered include cropping, mirroring, HSV colour space adjustments, and GN addition with two different variance levels (σ = 0.01 and σ = 0.1). The performance metrics show relatively small variations across different augmentation combinations. The 'Path' and 'Grassland' classes consistently maintain high accuracy and IoU scores, with minimal fluctuations across all augmentation scenarios. This suggests a robust classification for these terrain types, regardless of the augmentation strategy employed. The 'Other Vegetation' class shows slightly more variability, but the differences remain small. Similarly, the 'Car' and 'Unidentified' classes show only minor changes in performance metrics across different augmentation techniques. Interestingly, the addition of Gaussian noise, both at σ = 0.01 and σ = 0.1 levels, does not significantly impact the overall performance, with metrics remaining largely stable.

It is worth noting that the model had already a good performance on the original dataset, which likely explains the limited room for improvement in accuracy, IoU, and F1 metrics when applying data augmentation techniques. However, these

broad metrics alone do not fully capture the true value of the augmentation methods employed. The simple choice of one class over another does not reveal how confident the model is in its decision, or how far apart the chosen class is from the alternatives in terms of probability. This information can only be observed by analysing the output values of the SoftMax function.

The performance of the adopted deep learning model in terrain classification is shown in Figure 6. The top image displays the colour-coded terrain labels assigned to each pixel by the model. Below, a corresponding heatmap depicts the classification probability for each pixel, showing the model's confidence in its predictions, which was determined by taking the maximum value of the SoftMax output $\sigma(\boldsymbol{z})_i$. The heatmap is generated using four distinct intervals based on the maximum value of $\sigma(\boldsymbol{z})_i$: red indicates a confidence level below 68.3 %, orange represents values between 68.3 % and 95.4 %, yellow denotes the range from 95.4 % to 99.7 %, and green signifies confidence levels exceeding 99.7 %.

Two patterns emerge from the heatmap confidence image. Firstly, we observe a decrease in classification probability along the boundaries between different terrain types. As shown in [34], this could be attributed to aleatoric uncertainty, which is an inherent and irreducible uncertainty caused by sensor noise and partial observability. This type of uncertainty may potentially be reduced by improving sensor quality or increasing sensor resolution. Secondly, we notice areas with lower classification probabilities, which could arise from unfamiliar inputs not well-represented in the limited training dataset. This phenomenon is known as epistemic uncertainty, and it could be addressed and improved through data augmentation techniques. By applying methods such as geometric transformations, HSV changes, and Gaussian noise addition, we can enhance the model's performance in these areas by increasing the diversity and representation of training samples.

This trend of improved classification confidence through data augmentation is shown in Figure 7 and Figure 8. Figure 8 presents a series of image pairs for each training dataset, ranging from the original dataset to various combinations of augmentation techniques. Each pair consists of colour-coded terrain labels and corresponding classification probability heatmaps. This comparison shows the effect of different augmentation strategies on the model's confidence across the terrain.

Figure 7 provides a more quantitative analysis of this effect. This figure was generated by examining each image in the test

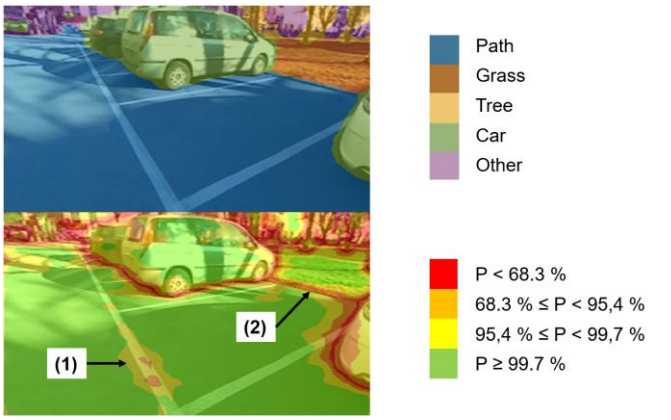

Figure 6. Comparison of pixel-wise terrain labels and associated classification probabilities. Top: Color-coded terrain labels assigned to each pixel by the deep learning model. Bottom: Heatmap representing the classification probability for each pixel. We can observe that the classification probability decreases: (1) at boundary zones between different labels—an intrinsic characteristic of the problem, (2) in areas where an adequate training dataset was not used—addressable through data augmentation.

sequence and calculating the percentage of pixels classified as "path" with a confidence level of 99.7 % or higher. We chose to focus on this specific class because the accurate identification of traversable paths is crucial for autonomous navigation and robot mobility in unstructured environments. These percentages were then used to create a boxplot, providing a statistical representation of the high-confidence pixel distribution across the entire test set. Table 3 shows the media percentage represented also in the boxplot.

The results presented in Figure 7 reveal a clear trend of increasing classification confidence with the application of more sophisticated data augmentation techniques. With only cropping applied, the median percentage of high-confidence path pixels is 75 %, serving as our baseline for comparison. The addition of mirroring and HSV transformations increases this median to 82 %, representing a substantial improvement in classification confidence. Further augmentation with GN yields even better results. Both combinations of cropping, mirroring, HSV, and GN with $\sigma = 0.01$ and cropping, mirroring, HSV, and GN with $\sigma = 0.1$ increase the median to 85 %. Interestingly, the combination of all techniques results in a slightly lower median of 84 %.

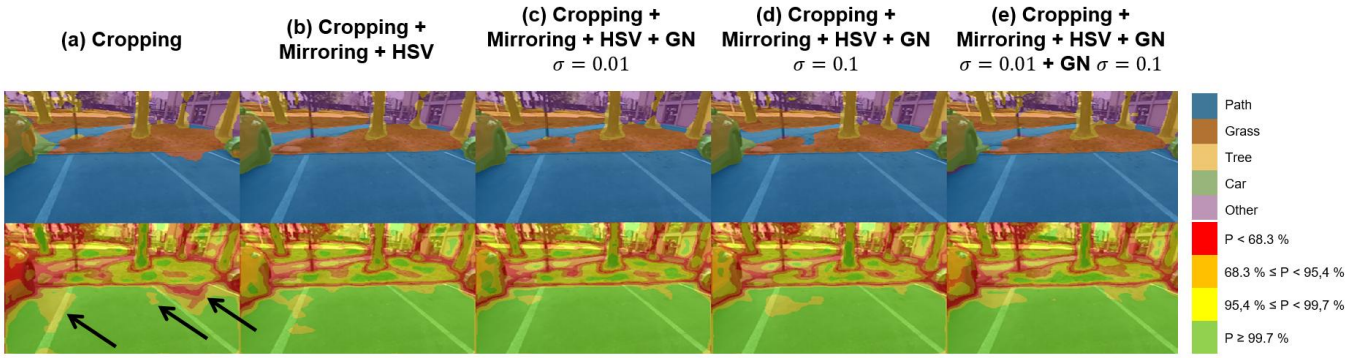

Figure 7. Comparison of pixel-wise terrain labels and classification probabilities across various training datasets. This figure presents a series of image pairs for each training dataset considered: (a) Original dataset; (b) Dataset with cropping augmentation, mirroring augmentation and HSV transformation augmentation (c) Dataset with combined augmentation techniques and Gaussian noise addition $\sigma = 0.01$; (d) Dataset with combined augmentation techniques and Gaussian noise addition $\sigma = 0.1$; (e) Dataset with combined augmentation techniques. For each pair, the top image shows color-coded terrain labels assigned to each pixel, while the bottom image displays a heatmap of classification probabilities.

Table 3. Median percentage of pixels classified as "path" with a confidence level greater than 99.7% in each image of the sequence for the various tested data augmentation techniques.

	Path
Cropping	75 %
Cropping + Mirroring + HSV	82 %
Cropping + Mirroring + HSV + GN σ = 0.01	85 %
Cropping + Mirroring + HSV + GN σ = 0.1	85 %
Cropping + Mirroring + HSV + GN σ = 0.01 + GN σ = 0.1	84 %

Our analysis suggests that, while traditional metrics show modest improvements, the application of data augmentation techniques appears to have a more pronounced effect on the model's prediction confidence. It is worth noting that the experiments suggest a potential drawback, as we observed a slight decrease in accuracy to 84 % when multiple augmentation methods were applied simultaneously. It is possible to note the impact of data augmentation on enhancing the model's confidence in terrain classification, particularly for critical categories such as traversable paths.

4. CONCLUSIONS

The research presented in this paper introduces an innovative architecture for terrain traversability analysis, specifically designed for UGV navigation in unstructured environments. Our approach leverages a supervised learning-based labelling system to map traversable terrain, utilising the DeepLabv3+ neural network architecture trained on a custom dataset collected from our specific testing location. A key aspect of our methodology was the use of data augmentation techniques to expand our dataset and enhance labelling accuracy. This approach proved to be crucial in addressing the challenges posed by the limited size of our initial dataset and the complexity of field terrain variations.

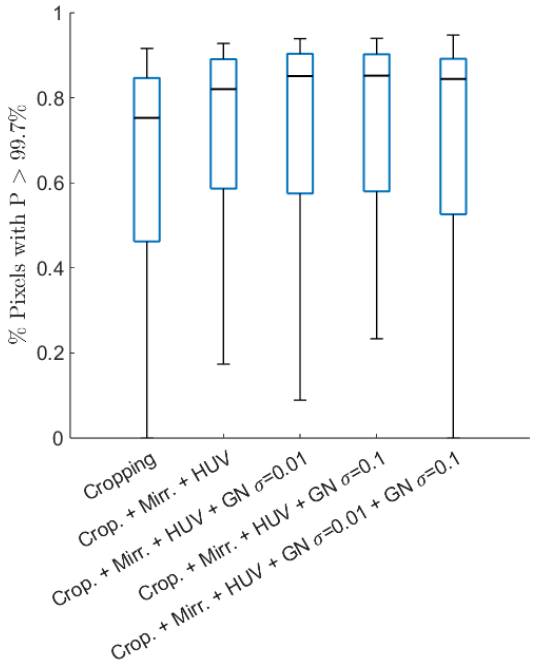

Figure 8. Distribution of pixels classified as "path" with confidence level exceeding 99.7%.

The results demonstrate the effectiveness of this approach in the context of UGV navigation. Evaluation using metrics such as accuracy and IoU shows good performance across terrain classes, including paths, grassland, and obstacles such as vehicles. Moreover, our analysis of classification probabilities highlighted the significant effect of data augmentation techniques—including cropping, mirroring, HSV transformations, and Gaussian noise addition—in enhancing the model's confidence in its predictions. To conclude, this paper suggest that carefully applied data augmentation can significantly enhance the robustness and reliability of terrain classification models, even when working with limited initial datasets.

Future directions for this research include expanding our dataset by collecting more diverse data to improve the model's performance across a wider range of terrain types and conditions. We also aim to explore more sophisticated data augmentation methods, such as Generative Adversarial Networks (GANs), for synthetic data creation. Additionally, we plan to experiment with different neural network architectures and ensemble methods to further improve classification accuracy and confidence. These efforts will build upon our current findings and contribute to advancing terrain classification for autonomous navigation in unstructured environments.

REFERENCES

[1] F. M. Favarò, N. Nader, S. O. Eurich, M. Tripp, N. Varadaraju, Examining accident reports involving autonomous vehicles in California. PLoS ONE, 2017, 12(9): e0184952.
DOI: 10.1371/journal.pone.0184952

[2] P. Papadakis, Terrain traversability analysis methods for unmanned ground vehicles: A survey. Engineering Applications of Artificial Intelligence, 26(4), 2013, pp. 1373-1385.
DOI: 10.1016/j.engappai.2013.01.006

[3] P. Marín-Plaza, J. Beltrán, A. Hussein, B. Musleh, D. Martín, A. de la Escalera, J. M. Armingol, Stereo Vision-based Local Occupancy Grid Map for Autonomous Navigation in ROS, Proc of the 11th Joint Conf. on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2016), Rome, Italy, 27 – 29 February 2016, volume 3, pp. 701-706.
DOI: 10.5220/0005787007010706

[4] A. Elfes, Using occupancy grids for mobile robot perception and navigation, Computer, vol. 22, June 1989, no. 6, pp. 46-57.
DOI: 10.1109/2.30720

[5] S. Thrun, W. Burgard, D. Fox, Probabilistic robotics, MIT press, 2005
DOI: 10.1162/artl.2008.14.2.227

[6] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, A. L. Yuille, DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs, in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 40, April 2018, no. 4, pp. 834-848.
DOI: 10.1109/TPAMI.2017.2699184

[7] X. Wu, N. Xu, J. Zhou, Terrain Classification Algorithm for Off–Road Vehicle Using 1D CNN–LSTM Network, in IEEE, 8th CAA Int. Conf. on Vehicular Control and Intelligence (CVCI), Chongqing, China, 25-27 October 2024, pp. 1-5.
DOI: 10.1109/CVCI63518.2024.10830195

[8] D. K. Shukla, K. Skonieczny, Simple texture descriptors for classifying monochrome planetary rover terrains, IEEE Int. Conf. on Robotics and Automation (ICRA), Singapore, 29 May-3 June 2017, pp. 5495-552.
DOI: 10.1109/ICRA.2017.7989647

[9] S. Quazi, S. M. Musa, Image Classification and Semantic Segmentation with Deep Learning, 6th IEEE Int. Conf. on Recent Advances and Innovations in Engineering (ICRAIE), Kedah, Malaysia, 1-3 December 2021, pp. 1-6.
DOI: 10.1109/ICRAIE52900.2021.9704014

[10] A. Garcia-Garcia, S. Orts-Escolano, S. Oprea, V. Villena-Martinez, P. Martinez-Gonzalez,J. Garcia-Rodriguez, A survey on deep learning techniques for image and video semantic segmentation, Applied Soft Computing, 70, 2018, pp. 41-65.
DOI: 10.1016/j.asoc.2018.05.018

[11] K. Alomar, H. I. Aysel, X. Cai, Data augmentation in classification and segmentation: A survey and new strategies, Journal of Imaging, 9(2), 2023, p. 46.
DOI: 10.3390/jimaging9020046

[12] L. Perez, The effectiveness of data augmentation in image classification using deep learning, arXiv preprint 2017 arXiv:1712.04621.
DOI: 10.48550/arXiv.1712.04621

[13] M. S. K. Sashank, V. S. Maddila, V. Boddu, Y. Radhika, Efficient deep learning based data augmentation techniques for enhanced learning on inadequate medical imaging data, Acta IMEKO, 11 (2022) 1, p. 6.
DOI: 10.21014/acta_imeko.v11i1.1226

[14] L. Wunsch, K. Anding, G. Polte, K. Liu, G. Notni, Data augmentation for solving industrial recognition tasks with underrepresented defect classes, Acta IMEKO, 12 (2023) 4, pp. 1-5.
DOI: 10.21014/actaimeko.v12i4.1320

[15] Y. Fu, J. Fan, S. Xing, Z. Wang, F. Jing, M. Tan, Image segmentation of cabin assembly scene based on improved RGB-D mask R-CNN, IEEE Transactions on Instrumentation and Measurement, 71, 2022, pp. 1-12.
DOI: 10.1109/TIM.2022.3145388

[16] L. Sun, J. Bockman, C. Sun, A framework for leveraging inter-image information in stereo images for enhanced semantic segmentation in autonomous driving, IEEE Transactions on Instrumentation and Measurement, 2023.
DOI: 10.1109/TIM.2023.3328708

[17] S. Chiodini, L. Torresin, M. Pertile,S. Debei, Evaluation of 3d cnn semantic mapping for rover navigation, in: IEEE 7th International Workshop on Metrology for AeroSpace (MetroAeroSpace), Pisa, Italy, 22-24 June 2020, pp. 32–36.
DOI: 10.1109/MetroAeroSpace48742.2020.9160157

[18] S. Chiodini, M. Pertile, S. Debei, Occupancy grid mapping for rover navigation based on semantic segmentation, Acta IMEKO 10 (2021) 4, pp. 155–161.
DOI: 10.21014/acta_imeko.v10i4.1144

[19] G. Polato, S. Chiodini,A. Valmorbida, (+ 3 more authors), Semantic Terrain Traversability Analysis Based on Deep Learning Aimed at Planetary Rover Navigation, Aerotec. Missili & Spazio, 104(3) 2024.
DOI: 10.1007/s42496-024-00238-0

[20] J. A. Placed, J. Strader, H. Carrillo, N. Atanasov, V. Indelman, L. Carlone, J. A. Castellanos, A survey on active simultaneous localization and mapping: State of the art and new frontiers, IEEE Transactions on Robotics, 2023, 39(3), pp. 1686-1705.
DOI: 10.1109/TRO.2023.3248510

[21] L.-C. -Chen, Y. Zhu, G. Papandreou, F. Schroff, H. Adam, Encoder-decoder with atrous separable convolution for semantic image segmentation, Proc. of the European Conf. on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018, pp. 801–818.
DOI: 10.1007/978-3-030-01234-2_49

[22] C. Guo, G. Pleiss, Y. Sun, K. Q. Weinberger, On calibration of modern neural networks, Int. conf. on machine learning, PMLR, July 2017, pp. 1321-1330.
DOI: 10.5555/3305381

[23] P. Y. Huang, W. T. Hsu, C. Y. Chiu, T. F. Wu, M. Sun, Efficient Uncertainty Estimation for Semantic Segmentation in Videos, in: V. Ferrari, M. Hebert, C. Sminchisescu, Y. Weiss (eds). Computer Vision – ECCV 2018. ECCV 2018. Lecture Notes in Computer Science, vol. 11205. Springer, Cham.
DOI: 10.1007/978-3-030-01246-5_32

[24] C Dechesne, P. Lassalle, S. Lefèvre, Bayesian u-net: Estimating uncertainty in semantic segmentation of earth observation images, Remote Sensing, 13, 2021 (19), p. 3836.
DOI: 10.3390/rs13193836

[25] C. Shorten, T. M. Khoshgoftaar, A survey on image data augmentation for deep learning. Journal of big data, 6, 2019 (1), pp. 1-48.
DOI: 10.1186/s40537-019-0197-0

[26] G. Divon, A. Tal, Viewpoint Estimation—Insights & Model, in: Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 252-268.
DOI: 10.48550/arXiv.1807.0131

[27] A. Jurio, M. Pagola, M Galar., C. Lopez-Molina, D. Paternain, A comparison study of different color spaces in clustering based image segmentation, in: Hüllermeier, E., Kruse, R., Hoffmann, F. (eds) Information Processing and Management of Uncertainty in Knowledge-Based Systems. Applications. IPMU 2010, Communications in Computer and Information Science, vol 81. Springer, Berlin, Heidelberg, 2010.
DOI: 10.1007/978-3-642-14058-7_55

[28] S. Boonprong, C. Cao, W. Chen,X. Ni, M. Xu, B. K. Acharya, The classification of noise-afflicted remotely sensed data using three machine-learning techniques: Effect of different levels and types of noise on accuracy, Isprs Int. J. Geo-Inf. 2018, 7, 274.
DOI: 10.3390/ijgi7070274

[29] Y. Liu, Y. Gao, W. Yin, An improved analysis of stochastic gradient descent with momentum, Advances in Neural Information Processing Systems, 33, 2020, pp. 18261-18271.
DOI: 10.48550/arXiv.2007.07989

[30] Q. H. Nguyen, H. B. Ly, L. S. Ho, N. Al-Ansari,H. V. Le, V. Q. Tran, B. T. Pham, Influence of data splitting on performance of machine learning models in prediction of shear strength of soil, Mathematical Problems in Engineering, 2021(1), p. 4832864.
DOI: 10.1155/2021/4832864

[31] J. Long, E. Shelhamer, T. Darrell, Fully convolutional networks for semantic segmentation, Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition, Boston, MA, USA, 7-12 June 2015, pp. 3431-3440.
DOI : 10.1109/CVPR.2015.7298965

[32] A. Garcia-Garcia, S. Orts-Escolano, S. Oprea, V. Villena-Martinez, J. Garcia-Rodriguez, A review on deep learning techniques applied to semantic segmentation, arXiv preprint, 2017, arXiv:1704.06857.
DOI: 10.48550/arXiv.1704.06857

[33] I. Z. Ibragimov, I. M. Afanasyev, Comparison of ROS-based visual SLAM methods in homogeneous indoor environment, 14th Workshop on Positioning, Navigation and Communications (WPNC), Bremen, Germany, 25-26 October 2017, pp. 1-6.
DOI: 10.1109/WPNC.2017.8250081

[34] S. Ancha, P. R. Osteen, N. Roy, Deep evidential uncertainty estimation for semantic segmentation under out-of-distribution obstacles, in: IEEE International Conference on Robotics and Automation (ICRA), Yokohama, Japan, 13-17 May 2024, pp. 6943-6951.
DOI: 10.1109/ICRA57147.2024.10611342