AI-Based Optimization of Roadside Mowing Operations in Austria

Roland Perko, Stefanie Onsori-Wechtitsch, Helmut Neuschmied,
Peter Schallauer, Katharina Hofer-Schmitz
JOANNEUM RESEARCH, Graz, Austria, `{firstname.lastname}@joanneum.at`

Michaela Stolz
biohelp, Vienna, Austria, `{firstname.lastname}@biohelp.at`

Abstract

Roadside vegetation management is vital for traffic safety, efficiency, and biodiversity. Conventional mowing relies on routine schedules and manual inspections, limiting route optimization and adaptation to changing vegetation growth. To address the challenges of roadside maintenance, we developed MeadowLevelSeg, a deep learning approach that employs Mask2Former to map meadow heights into precise 5 cm classes. Around 800 high-resolution roadside images were recorded and annotated. Performance is evaluated using a novel Distance-Aware Accuracy metric, which takes the ordinal nature of height classes into account. Initial results demonstrate that the model effectively identifies different meadow heights and high-growth zones, achieving a mean absolute error of less than 7 cm using monocular images. This provides a robust basis for automated maintenance scheduling.

1 Introduction

Roadside vegetation management is essential for traffic safety, operational efficiency, and biodiversity along public roads. Traditionally, roadside mowing is labor-intensive, guided by weekly lists and routine checks, with limited capacity to optimize routes or adapt to dynamic growth patterns. This necessitates balancing economic and environmental goals, characterizing roadside maintenance as a complex multi-objective optimization problem [8]. This work aims to leverage AI-based computer vision for estimating meadow height to optimize mowing schedules and routes while balancing road safety and ecological goals. The visibility of safety equipment close to the road (e.g., the reflectors of guiding posts) is given priority. Furthermore, overgrown pathways that pose risks to children and cyclists are systematically identified and reported at an early stage. The research is also motivated by the *Federal Law on the Procurement and Use of Clean Road Maintenance Vehicles* [1], which requires road maintenance authorities to transition to emission-free fleets by 2030. Since those vehicles are limited in their maximum range, optimized route planning is mandatory. Although computer vision is well-established in agriculture, its application to measuring the height of roadside meadows remains largely unexplored. While existing Digital Twin approaches (e.g., [9]) address the safety risks of vegetation overgrowth, they primarily focus on detecting overhanging trees rather than providing fine-grained classification of meadow heights. A similar approach was followed by [7], who assess ground conditions, such as soil moisture and trafficability, directly on a mower. This work aspires to provide road maintenance authorities with actionable insights into when and where mowing is needed, improving transparency, reducing unnecessary interventions, and supporting sustainable management. Compared to costly LiDAR systems, our camera-based MeadowLevelSeg offers a more cost-effective, easily integratable alternative. This study presents initial results on the semantic segmentation of meadow height classes, which serves as a prerequisite for GIS-based mowing optimization. For illustration, Figure 1 (left) shows a mowing vehicle in operation.

The Third Austrian Symposium on AI, Robotics, and Vision (AIROV26).

Figure 1: Road maintenance service vehicle from the province of Styria during mowing activity (left). Statistical breakdown of the annotated dataset (right). The histograms show a comparison between total annotated area (top) and polygon counts (bottom) across 5 cm height increments.

2 Data Collection

A sensor platform was constructed which combines high-end imaging (dual Basler acA4096-30uc industrial cameras), low-cost monocular imaging (GoPro Hero 13), edge AI processing (Intel i7, NVIDIA RTX 3070 GPU), high-accuracy EGNSS positioning, and synchronized custom hardware triggers. The system was installed in a road maintenance vehicle and its components were waterproofed and thoroughly calibrated to ensure radiometric fidelity. From April to October 2025, the sensor platform captured roadside imagery along predefined routes around Feldbach in Styria. A GIS-based planning tool was used to select routes and features, with the aim of targeting different vegetation conditions, slopes, bushes, trees, and pollinator habitats. Annotations were performed using the *Computer Vision Annotation Tool (CVAT)* [4], focusing on meadow height classes from 5 to 80 cm with increments of 5 cm. All relevant roadside strips (50 cm wide) next to roads and paths were annotated for a given dataset, which currently comprises approximately 800 images. Figure 1 (right) shows the distribution of meadow classes in the annotated data, with the percentage share of the whole annotated area in image space on the top and the annotated polygons on the bottom. While the distribution of annotated pixels suggests a relatively balanced dataset across almost all classes, a comparison with the number of annotated polygons reveals a significant under-representation of the classes 30 cm to 80 cm. This highlights the need for targeted data collection during the next vegetation period, in order to improve the diversity of the training data.

3 Methodology

The primary goal of the proposed MeadowLevelSeg framework is to estimate meadow heights along roadside strips. For this specific semantic segmentation task, we adopted Mask2Former [3] architecture, as its mask-classification approach outperforms traditional per-pixel models like UPerNet [10] in capturing the fine, irregular boundaries of vegetation. Furthermore, Mask2Former provides an optimal balance between inference speed and accuracy, which is crucial for the high-resolution processing demanded by roadside maintenance tasks. The dataset consists of 833 high-resolution roadside monocular images ($4,112 \times 2,176$ pixels) which were annotated in CVAT and exported as segmentation masks. This set was randomly split into 617 training, 152 validation, and 64 test images. The architecture was implemented within the OpenMMLab framework [2], using a Swin Transformer backbone [5] initialized with pretrained ADE20K weights [11] before being fine-tuned on our domain-specific roadside data. To better align the model with the specific application requirements, the training strategy was adapted by modifying the loss formulation and optimization process (also cf. [6]). Even minor height class deviations have a significant impact on the standard metric, which is why distance-aware accuracy (DistAwareAcc) was introduced as a more representative performance metric. This semantic segmentation metric measures accuracy while accounting for the semantic distance between classes. Instead of treating all misclassifications equally, it penalizes predictions proportionally to how different the predicted height class is from the ground-truth class.

Distance-Aware Accuracy. Let K be the total number of classes, where classes 1 to $K-1$ represent height classes and class 0 represents the background class. $C \in \mathbb{N}^{K \times K}$ be the confusion matrix, where C_{ij} denotes the number of pixels of ground-truth class i predicted as class j. We define a class-similarity matrix $W \in [0,1]^{K \times K}$ that encodes the semantic proximity between categories. The weights W_{ij} are determined by the ordinal distance between class indices for meadow, while ensuring no tolerance for background misclassifications:

$$W_{ij} = \begin{cases} 1.0 & \text{if } i = j \\ \max(0, 1.0 - 0.1 \cdot |i-j|) & \text{if both } i, j \text{ are height classes and } i \neq j \\ 0.0 & \text{if either } i \text{ or } j \text{ is the background class and } i \neq j \end{cases} \tag{1}$$

By setting $W_{ij} = 0.0$ for any mismatch involving the background, the model maintains a strict boundary between meadow and non-meadow areas. For meadow classes, the weight decreases in steps of 0.1 for every discrete height interval (5cm steps). The per-class accuracy is calculated as:

$$\mathrm{DistAwareAcc}_i = \frac{\sum_{j=1}^{K} C_{ij}\, W_{ij}}{\sum_{j=1}^{K} C_{ij}} \tag{2}$$

The overall $\mathrm{DistAwareAcc}$ is the mean across all classes and is used within the loss function to reduce penalization of semantically similar height predictions during training.

4 Results

Intersection over Union (IoU) was utilized as the primary metric for validating the semantic segmentation of roadside meadow height regions. For the test dataset, the model achieved an average IoU of 17.3% across all classes. However, it should be noted that a perfect overlap is intrinsically limited by two factors. Firstly, there is the inherent ambiguity in annotating non-uniform meadow heights. Secondly, there are the perspective-related challenges of defining polygon boundaries at a fine level. Given the inherent uncertainties associated with annotation, the results indicate a high level of reliability for operational use. As demonstrated in Table 1, a comprehensive analysis of the performance of each class is provided. The low scores, particularly the score of 0 for classes 55 cm and 80 cm, can be attributed to the strong under-representation of these classes in both the training and test datasets. With a limited number of ground-truth pixels available, the model predicts very few instances, resulting in minimal to zero overlap. A similar lack of intersection is observed for the 35 cm, 60 cm, and 70 cm classes. In the case of the 80 cm class specifically, the model failed to detect any corresponding areas. As there were no positive predictions (true or false), the precision remains undefined (NaN). The $\mathrm{DistAwareAcc}$ metric, which is more relevant for the absolute meadow height estimation, yields significantly higher values. It indicates that while the model is not always capable of accurately identifying the exact height, it frequently assigns a neighboring height class instead of the exact class. This is further demonstrated by the qualitative results, presented in Figure 3. The detected classes are overlaid on the input image. Compared to manual annotations, which often group larger areas into a single height, the model often delivers more precise results with several different height classes. It captures details that were either simplified or overlooked during manual annotations.

Mean Absolute Error (MAE). To quantify the prediction quality, the MAE was calculated across all meadow pixels on the test set. By converting the class indices back to the corresponding heights in centimeters, it was determined that the model's predictions deviate from the actual height by only 6.57 cm on average. This minor discrepancy in spatial height confirms that, although the exact 5 cm category is not achieved, the anticipated trend for operational planning remains highly precise.

Table 1: Segmentation results per meadow height class. While IoU, Precision, and Recall reflect the difficulty of exact classification, the $\mathrm{DistAwareAcc}$ shows the capability for operational planning.

Class [cm]	5	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80
IoU [%]	39.0	11.5	23.2	18.0	27.9	15.0	0.5	11.8	5.6	18.4	0.0	5.0	13.8	4.7	8.5	0.0
Precision [%]	56.5	13.7	33.5	28.0	38.4	21.4	5.4	14.8	12.8	30.4	0.0	14.4	17.9	5.9	47.0	NaN
Recall [%]	55.7	42.4	42.9	33.5	50.6	33.2	0.5	36.9	9.1	31.7	0.0	7.2	37.4	19.5	9.3	0.0
DistAwareAcc [%]	82.9	84.1	87.6	87.4	85.0	81.9	76.7	81.7	77.9	82.3	67.0	81.1	70.4	69.9	60.6	56.4

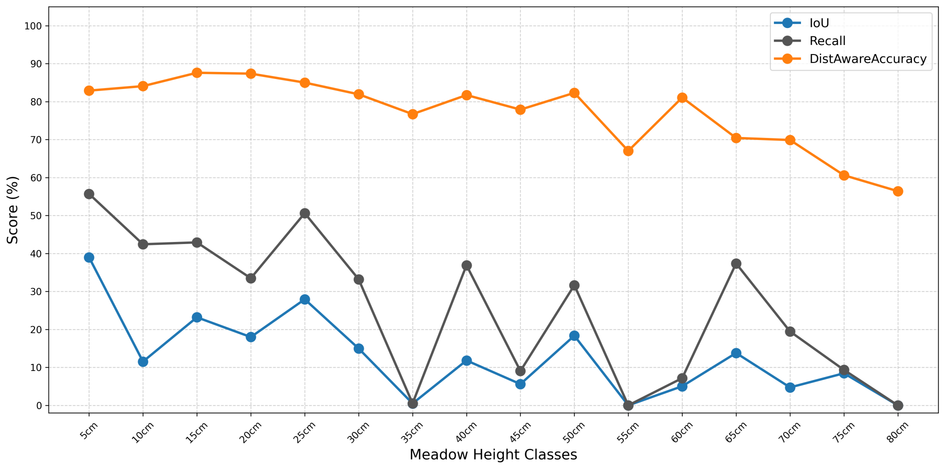

Figure 2: Detection performance of Meadow Heights when considering neighboring classes and their similarities.

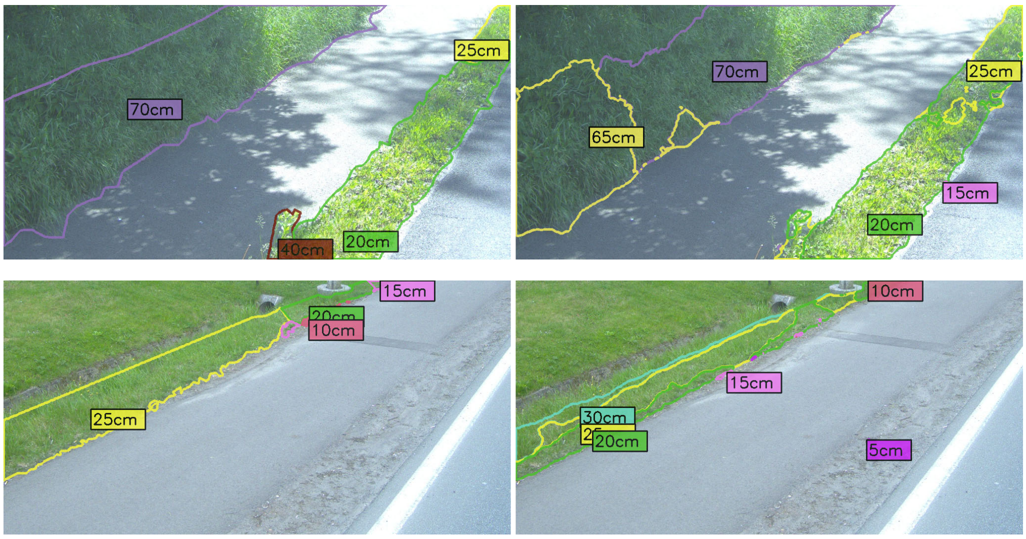

Figure 3: Exemplary results of semantic segmentation into meadow height classes – ground-truth annotated meadow height classes (left), predicted meadow height classes (right). Both examples show cycling paths directly adjacent to the road.

5 Conclusions

The presented work demonstrates feasibility for scalable AI-guided meadow management on Austrian roads. The results indicate promising generalization behavior and reasonable agreement between predictions and ground truth, where the error of the meadow height estimation was below 7 cm. However, under-representation of high meadow classes suggests areas for improvement in future annotation efforts and model refinement, which are planned for the upcoming vegetation period.
The next steps will focus on (1) increasing annotation density across all height classes and diverse geographic routes to improve model generalization, (2) develop an automated mowing-planner for optimized routing based on meadow height classification of an entire road network, and (3) design and implement a proof-of-concept demonstrator, allowing the end users to evaluate the whole system. The groundwork is laid for a collaborative, evidence-based management strategy that balances safety requirements, cost-efficiency, and ecological biodiversity goals.

Acknowledgments and Disclosure of Funding

This research was partly funded by the AI for Green program through the project SmartMowAI (FFG project number 910271).

References

[1] Bundesministerium für Justiz. BGBl. I Nr. 163/2021: Bundesgesetz über die Beschaffung und den Einsatz sauberer Straßenfahrzeuge - Federal Act on the Procurement and Use of Clean Road Vehicles (Road Vehicle Procurement Act). https://www.ris.bka.gv.at/eli/bgbl/I/2021/163, 2021.

[2] Kai Chen, Jiaqi Wang, Jiangmiao Pang, Yuhang Cao, Yu Xiong, Xiaoxiao Li, Shuyang Sun, Wansen Feng, Ziwei Liu, Jiarui Xu, Zheng Zhang, Dazhi Cheng, Chenchen Zhu, Tianheng Cheng, Qijie Zhao Zhao, Buyu Li, Xin Lu, Rui Zhu, Yue Wu, Jifeng Dai, Jingdong Wang, Jianping Shi, Wanli Ouyang, Chen Change Loy, and Dahua Lin. MMDetection: Open MMLab detection toolbox and benchmark. *arXiv preprint arXiv:1906.07155*, 2019.

[3] Bowen Cheng, Ishan Misra, Alexander G Schwing, Alexander Kirillov, and Rohit Girdhar. Masked-attention mask transformer for universal image segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1290–1299, 2022.

[4] CVAT Contributors. Computer Vision Annotation Tool (cvat). https://github.com/cvat-ai/cvat, 2024. Version 2.48.0.

[5] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin transformer: Hierarchical vision transformer using shifted windows. In *International Conference on Computer Vision*, pages 10012–10022, 2021.

[6] Jun Ma, Jianan Chen, Matthew Ng, Rui Huang, Yu Li, Chen Li, Xiaoping Yang, and Anne L Martel. Loss odyssey in medical image segmentation. *Medical Image Analysis*, 71:102035, 2021.

[7] Christoph Manss, Viktor Martel, and Roman Weisgerber. Assessment of ground conditions in grassland on a mower with artificial intelligence. In *GIL-Jahrestagung, Biodiversität fördern durch digitale Landwirtschaft*, pages 335–340, 2024.

[8] Adriana-Simona Mihaita, Brunelle Marche, Mauricio Camargo, Iman Rahimi, and Christophe Bachmann. Multi-objective modelling of a roadside mowing problem: A case study in France. In *IEEE International Conference on Engineering, Technology and Innovation & International Association For Management of Technology, Joint Conference*, pages 1–8, 2022.

[9] Varun Kumar Reja, Diana Davletshina, Mengtian Yin, Ran Wei, Quentin Felix Adam, Ioannis Brilakis, and Federico Perrotta. A digital twin based approach to control overgrowth of roadside vegetation. In *Proceedings of the International Symposium on Automation and Robotics in Construction*, pages 1–8, 2024.

[10] Tete Xiao, Yingcheng Liu, Bolei Zhou, Yuning Jiang, and Jian Sun. Unified perceptual parsing for scene understanding. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 36–51, 2018.

[11] Bolei Zhou, Hang Zhao, Xavier Puig, Tete Xiao, Sanja Fidler, Adela Barriuso, and Antonio Torralba. Semantic understanding of scenes through the ADE20K dataset. *International Journal of Computer Vision*, 127(3):302–321, 2019.