Labelling point clouds in VR

Cyprien Raymi Fol[1], Arnadi Murtiyoso[1,2], Gabriele Mazzacca[3,4], Thomas Saint-André[2],
Fabio Remondino[3] and Verena Christiane Griess[1]

[1] Forest Resources Management, Institute of Terrestrial Ecosystems, Department of Environmental Systems Science, ETH Zurich, Switzerland – (cyprien.fol, arnadi.murtiyoso, verena.griess)@usys.ethz.ch

[2] Université de Strasbourg, INSA Strasbourg, CNRS, ICube Laboratory UMR 7357, Photogrammetry and Geomatics Group, 67000 Strasbourg, France – (arnadi.murtiyoso, thomas.saint-andre)@insa-strasbourg.fr

[3] 3D Optical Metrology (3DOM) unit, Bruno Kessler Foundation (FBK), Trento, Italy – (gmazzacca, remondino)@fbk.eu

[4] University of Udine, Department of Mathematics, Computer Science and Physics, Udine, Italy - mazzacca.gabriele@spes.uniud.it

Keywords: Point Cloud, Virtual Reality, Training Set, Cultural Heritage, Forestry.

Abstract

Recent advancements in Virtual Reality (VR) technology have extended its applications beyond entertainment, offering promising tools for professional fields such as 3D data annotation. This paper explores the use of VR for labelling 3D point clouds in forestry and cultural heritage datasets. We employ Labelling Flora, an open-source VR annotation tool, to re-label three existing cultural heritage and one forestry datasets and assess the effectiveness of VR-based annotations in training machine learning models. By comparing the accuracy, precision, and F1-score of inference models trained with VR-generated labels to those trained with traditional desktop labelling methods, we evaluate the potential of VR to streamline labour-intensive annotation tasks. Our results indicate that VR enables intuitive 3D segmentation, even for individuals without technical expertise, particularly for very complex scenes, improving labelling efficiency and contributing to the overall automation of complex datasets. This study highlights therefore the potential of VR to enhance other workflows and make complex tasks more accessible to domain experts who may not be familiar with 3D data thus refining data accuracy and reliability.

1 Introduction

Virtual Reality (VR) has seen remarkable advancements in the past decade, driven by the development of new headsets and increased competition in the consumer market. While VR is gaining popularity in the leisure industry due to its immersive and engaging experiences, its applications have expanded to professional fields such as medical training and architectural visualization, where it enhances spatial awareness and precision (Suh et al., 2018).

Recent research has explored the potential of VR for visualizing (Kharroubi et al., 2019; Zhao et al., 2019) and annotating 3D point clouds (Fol et al., 2022), yielding promising results. Researchers have also conducted user studies to assess the benefits of VR-based labelling tools compared to traditional 2D desktop-based software. For example, Franzluebbers et al. (2022) developed and tested a VR-based labelling tool designed for plants' 3D scans, with users reporting enhanced accuracy, enjoyment, and a greater sense of precision. Similarly, Venn & Mills (2023) compared VR-based labelling with traditional desktop methods, revealing statistically significant improvements in performance and user perception.

Despite these promising developments, labelling tasks are primarily undertaken to create training sets that automate the labour-intensive process of annotating 3D data. However, no research to date has specifically examined the effectiveness of VR-based labels for training machine learning (ML) models. The key question remains: **Are VR-based annotations accurate enough to train reliable machine learning models?**

To fill this gap, this paper employs Labelling Flora (Fol et al., 2024), an open-source VR-based annotation tool, to re-label various existing datasets from the forestry, and cultural heritage sectors (Figure 1). The re-labelled datasets will be used to train machine learning models, and the resulting segmentation performance will be evaluated in terms of accuracy, precision, recall, intersection over union, and F1-scores. This approach will determine whether models trained with VR-based labels can achieve results comparable to those trained with conventional desktop-based labels. Importantly, this study aims to bridge a gap between domain experts - who may lack the technical skills to operate complex 3D modelling software - and computer scientists, who may not possess the domain-specific knowledge required for accurate labelling. By leveraging the intuitive and immersive nature of VR, the goal is to make machine learning applications more accessible to professionals. The study seeks to determine if VR may support accurate 3D classification results using a more interactive method for annotation/labelling which is potentially also faster than a desktop-based approach.

The experiment's design and tests are primarily inspired by the work of Grilli & Remondino (2020) and Fol et al. (2024) while for the 3D classification a popular machine learning algorithm is used.

2 Materials and Methods

2.1 Datasets

For this study, existing datasets from the literature are used: a first from the forestry sector (Fol et al., 2022), consisting of single-tree stems and three datasets from the cultural heritage sector. All the datasets used in this study (Figure 2) have been labelled manually for machine learning purposes and are thus available for download (see Appendix).

Cultural heritage dataset: three datasets were used in this experiment to illustrate different types of architecture with different levels of complexity. The three used datasets are:

(a) (b)

Figure 1. The two labelling processes: (a) desktop-based and (b) VR based using Labelling Flora (Fol et al., 2024).

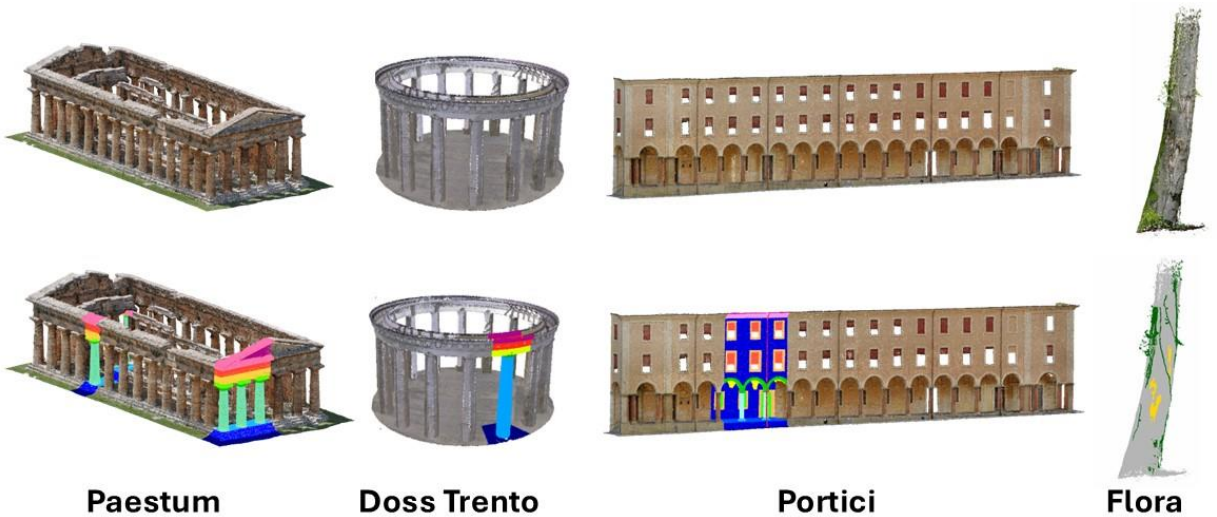

Figure 2. The datasets used in this paper: original point cloud (top row) and labelled parts used for the training process (bottom row).

1. "Paestum" dataset: acquired using a terrestrial laser scanner (TLS), it consists of the following classes: Grass, Crepidoma, Pavement, Shaft, Echinus, Abacus, Architrave, Frieze, Cornice, and Tympanum (Grilli & Remondino, 2019).
2. "Doss Trento" dataset: a TLS point cloud of the Cesare Battisti Mausoleum monument in Trento, consisting of six classes: Floor, Shaft, Capital, Architrave, Frieze, and Cornice. It is part of the NeRFBK dataset (Yan et al., 2023).
3. "Portici" dataset: it is a photogrammetric point cloud of a façade from the city of Bologna and it consists of 13 classes: Road, Pavement, Capital, Curtain, Façade, Arch, Molding, Columns base, Vault, Drainpipe, Shaft, Window/door and Cornice. It is part of the 3DOM Semantic Façade dataset (Stathopoulou et al., 2021).

Forestry dataset: it consists of 17 point clouds of tree stems within the "Flora" dataset (Fol et al., 2022), generated using close-range photogrammetry (CRP). The tree stems were labelled to extract biodiversity indicators along the stems, using both VR- and desktop-based methods. The tree stem is split into five distinct classes: Non-TreM, Tree injuries and exposed wood, Cavities, Epiphytic and epixylic structures and Excrescences and fruiting bodies of saproxylic fungi. These classes constitute the tree-related microhabitats (TreMs) (Fol et al., 2022; Rehush et al., 2018).

3 Methods

Figure 3 illustrates the pipeline used in this study to process both the forestry and cultural heritage datasets. The workflow starts with re-labelling the training datasets using the Labelling Flora application, developed within the Unity Game Engine (Fol et al., 2024). It is worth noting that all the datasets used in this experiment have been labelled manually using a desktop-based segmentation tool.

For the VR labelling process, an Oculus CV1 VR headset was employed. While Labelling Flora was developed for the HTC Vive VR headset (Fol et al., 2024), it has however demonstrated practically seamless cross-hardware implementation as shown by the experiment conducted in this work. Once the datasets are re-labelled, the Random Forest for Point Cloud Classification (RF4PCC) method (Grilli & Remondino, 2020) is used to perform the semantic segmentation prediction on the rest of the dataset.

Finally, the VR-based results are compared to the test sets generated using the desktop-based methods. For the comparison, common machine learning metrics are used, namely: average intersection over union (IoU), overall accuracy (oacc), average precision (P), average recall (R) and average F1 score (F1).

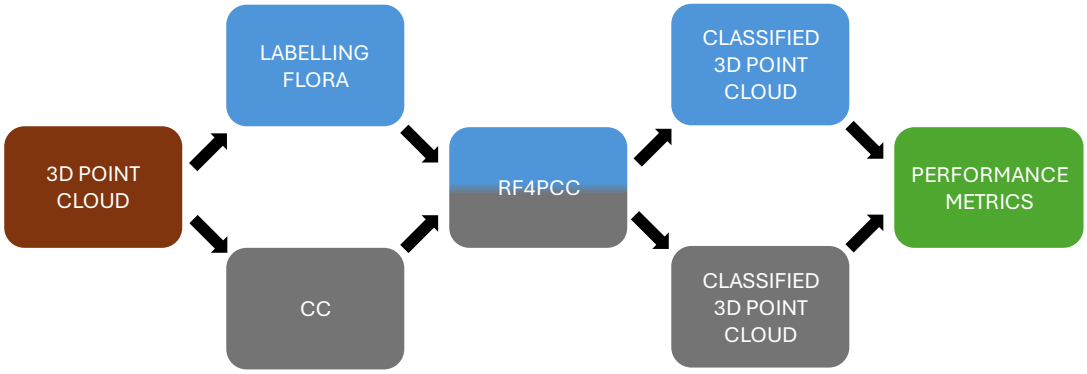

Figure 3. Workflow for evaluating VR-based labelling methods (blue boxes) versus desktop-based methods (grey boxes).

3.1 Adjusting Labelling Flora VR tool

To achieve proper labelling with any type of dataset, several modifications were made to the Labelling Flora VR application (Fol et al., 2024). In the original tool, the point cloud is inserted into a bounding box that encapsulates all points of the considered cloud. By moving the centre of this box, the point cloud itself can be repositioned. For the heritage datasets the box was off-centre and it needed to be shifted more towards the centre of the VR zone (limited to around 2x2 metres in this scenario) otherwise the limited space would interfere during labelling operations. Moreover, since the heritage datasets include objects of a bigger scale compared to the Flora datasets, scaling functionalities were added to enable better manipulation during the labelling process. To further improve navigation and labelling efficiency for larger point clouds in Labelling Flora, additional movement features were reconsidered.

Currently, the application does not support translating the point cloud along the XY plane—such as by using controller joysticks, a common feature in video games. This design choice was intended to encourage users to physically move around the point cloud, simulating real-world interaction and reducing the risk of cyber-sickness, especially for users less familiar with VR. However, this restriction proves less effective for large-scale datasets, such as those found in cultural heritage collections that feature buildings or monuments. To address this, future updates may introduce scaling adjustments alongside new navigation options, including 'fly-through' movement or XY plane translation, to enhance usability and flexibility with these larger datasets.

4 Results and Discussions

4.1 VR labelling process

For the Paestum and Doss Trento datasets, labelling was more complex due to the presence of corners and surfaces that are challenging to evaluate on a 2D desktop screen. This is, however, where VR shows advantages, allowing for multiple clear viewpoints of the object in a faster way. Thanks to the appearance of precise layers in real time within the Labelling Flora interface, a two-step yet intuitive labelling process was also possible: the first being a rough annotation followed by a refinement afterwards.

The Portici dataset presented a different challenge due to the largest number of classes (13) thus requiring a longer time for annotations. Its verticality also revealed a limitation to the VR labelling process, and it may not be user-friendly for those with mobility issues.

Overall, the labelling process is made far less complicated as the user should see the point cloud instead of rotating it, making cross-sections, segmenting, etc., in a 2D space. The user therefore does not need to be highly knowledgeable in 3D data to label a point cloud. Familiarity with the semantic layers and the elements to be inserted into them is sufficient. Additionally, being able to display the desktop screen at the bottom of the view in a small window made the labelling process more efficient, as the user did not have to constantly switch between the real desktop trying to remember all the layers. The process thus presents a form of gamification for the point cloud labelling process.

In terms of labelling time, the VR approach (Table 1) resulted in a similar time to a classical desktop-based approach except for the forestry dataset, where VR was faster due to the complexity and details of the scene, enabling at the same time the user to interact and quickly rectify any possible errors. While an exact numerical comparison of labelling time is not yet available, Fol et al. (2024) reported a reduction of around 56% compared to the time required by labelling using the screen/desktop. This value was achieved only for the Flora dataset.

	Flora	Paestum	Doss Trento	Portici
Labelling time [h]	3:50	0:50	0:18	1:30

Table 1. Labelling time required to perform the annotation tasks using Labelling Flora in VR.

4.2 Comparison of machine learning metrics

Starting from the VR- and desktop-based (Figure 4), the 3D classification was performed (Figure 5): it can be observed that classifications based on both VR and desktop labels are visually similar. The inference model performed well in distinguishing non-TreM points but struggled with accurately defining the borders of the TreM subcategories. These findings suggest that VR-based labelling could be a viable alternative to traditional methods. Moreover, the challenges encountered by the ML model in segmenting the forestry dataset highlight VR's potential to enhance the accuracy of machine learning outputs.

A comparison in percentage points is shown in Table 2, where the VR labelling managed to outperform the traditional method in almost all metrics. Indeed, in terms of recall, overall accuracy, and IoU, the improvement averages around 20% compared to traditional desktop-based labelling (see Figure 5). While this can be attributed to the advantage of VR in interacting with the highly heterogeneous environment present in forest scenes, this should nonetheless be considered only as an initial indication.

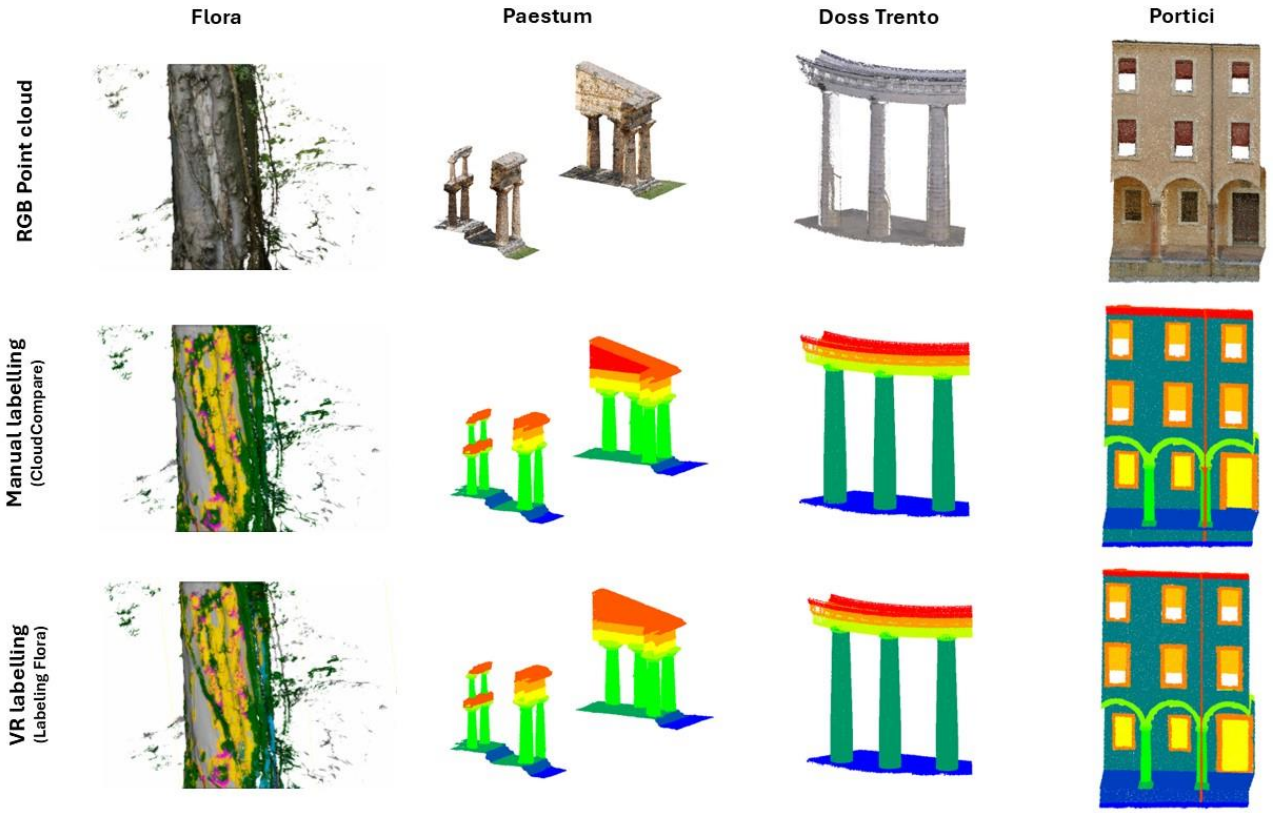

Figure 4. Visual comparisons of the labelling results for the four datasets for the purposes of creating training data (desktop-based CloudCompare and VR-based Labeling Flora).

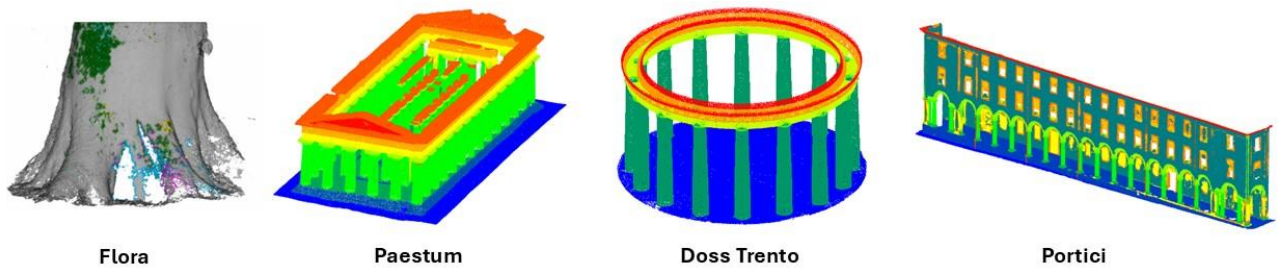

Figure 5. Results of the prediction based on RF4PCC using the training datasets generated by VR labelling as inputs.

More tests using forest scenes must be conducted before a definitive conclusion may be drawn.

A comparison in percentage points is shown in Table 2 and Figure 6, where the VR labelling managed to outperform the traditional method in almost all metrics in particularly for the forestry dataset. Indeed, in terms of recall, overall accuracy, and IoU, the improvement averages around 20% compared to traditional desktop-based labelling (see Figure 5). While this can be attributed to the advantage of VR in interacting with the highly heterogeneous environment present in forest scenes, this should nonetheless be considered only as an initial indication. More tests using forest scenes must be conducted before a definitive conclusion may be drawn.

A similar trend – but not so significative - can be observed in the three cultural heritage datasets, though at a far less convincing rate. Indeed, although the metrics for the VR datasets are slightly better than the desktop ones, it is statistically insufficient to draw any conclusions at this point regarding the improvement in quality. What may be safely ascertained, however, is that the VR method managed to attain a quality comparable to desktop means at least concerning datasets for machine learning purposes.

It is also interesting to note that for architectural datasets, the RF4PCC algorithm managed to score very good metrics, and the VR labelling does not significantly increase them. This contrasts with the forest dataset, where the prediction quality is lower, but the VR labelling process managed to improve the prediction quality to a greater degree. Much of this can be explained by the nature of the scenes themselves. Indeed, in the three cultural heritage datasets, classes were based on architectural elements, which are man-made and therefore more clear-cut.

On the other hand, in the forest dataset, the classes divided by the type of TreM do not translate into easily delineated objects. In several cases, interpretation by an expert is required, thus making it difficult to maintain a standardised classification. This also adds to the complexity of the heterogeneity and unordered nature of forest scenes in performing labelling and, by extension, machine learning predictions.

We argue that in these complex cases, the VR method is even more useful as it allows domain experts (e.g., foresters, biodiversity scientists) who are unfamiliar with 3D labelling techniques or the concept of point clouds in general to use it with greater ease. This is due to the intuitive nature of the VR application, which was designed to follow basic painting

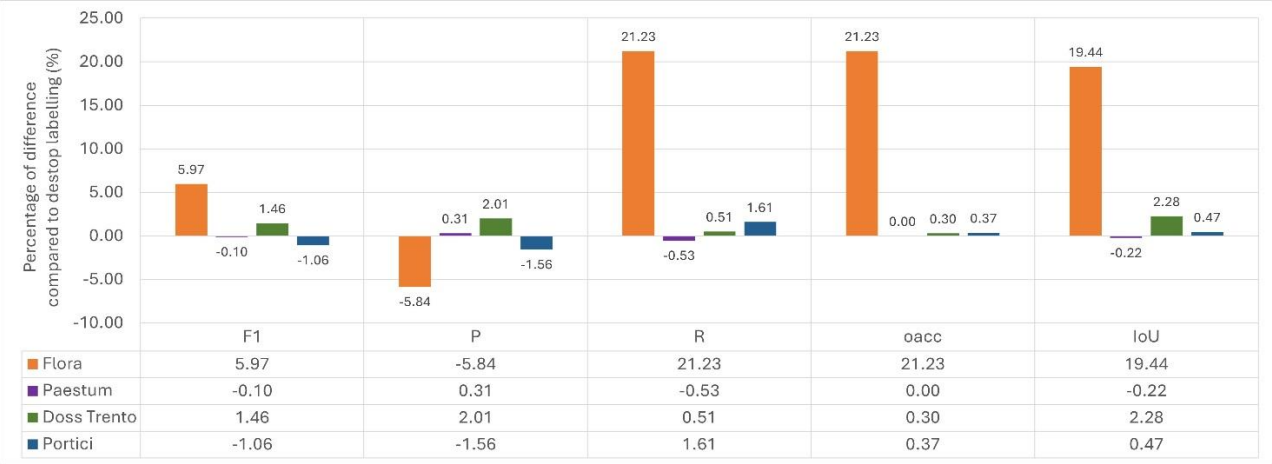

Figure 6. Difference (in percentage points) of machine learning metrics generated using the VR-labelled training data with respect to a traditional desktop-labelled datasets.

		F1	P	R	OverAcc	IoU
Flora	Desktop	0.787	0.891	0.730	0.730	0.360
	VR	0.834	0.839	0.885	0.885	0.430
Paestum	Desktop	0.953	0.956	0.951	0.967	0.912
	VR	0.952	0.959	0.946	0.967	0.910
Doss Trento	Desktop	0.957	0.944	0.974	0.984	0.923
	VR	0.971	0.963	0.979	0.987	0.944
Portici	Desktop	0.752	0.834	0.747	0.811	0.635
	VR	0.744	0.821	0.759	0.814	0.638

Table 2. Comparison of inference metrics (average F1, P, R and IoU values across classes).

movements. This, in turn, highlights the necessity of presenting an easy-to-use and ergonomic design for the user interface to complement the intuitive nature of the VR system.

5 Conclusions

The paper evaluated the applicability of VR-based labelling across datasets through specific use-case scenarios. Our primary goal is to demonstrate that VR enables individuals, including those without technical expertise, to support 3D segmentation tasks in an intuitive way by facilitating annotations of 3D point clouds. By integrating VR with machine/deep learning methods demanding annotated data, the labels generated can facilitate the segmentation of larger and complex datasets. Given that traditional labelling is a time-consuming task for scientists and researchers, VR presents a promising solution, allowing domain experts to label data more efficiently. Furthermore, VR can also be utilised to refine existing labels and retrain inference models, potentially enhancing overall accuracy.

This paper has shown that the VR labelling method is capable of producing annotated point clouds within a generally faster timeframe while achieving a classification quality comparable to results created using traditional desktop-based methods. Furthermore, the proposed VR method demonstrated its advantage more clearly when working with complex scenes such as forests. While there is a possibility that the inclusion of this immersive and intuitive element could improve the quality of ML-based predictions, particularly in more complex scenarios such as forests and trees, further tests are needed to draw statistically significant conclusions in other sectors. However, it has been demonstrated that the use of VR in labelling 3D datasets is a reliable method that is easily repeatable even by those who have only started working with point clouds. The proposed method is therefore accurate enough to be reliably used for machine learning annotation purposes. Finally, a proper user study performed across different datasets should be conducted in order to be able to draw statistically sound conclusions regarding the use of VR in 3D data annotation. This is despite the fact that the proposed method showed promise in its flexibility and user friendliness.

Several potential improvements to the Labelling Flora application became evident during the course of the experiments, including:

- Enhancements for visual comfort during long labelling sessions, e.g., increasing colour contrast and lowering the screen's gamma.
- Implementation of variable levels of detail for the input point cloud, using a culling system (for distant points) and an octree structure (for close points) to allow users to zoom in on the point cloud.
- Improving physical comfort for the operator by implementing third-person XY + Z movement to allow users to navigate virtually without having to move physically.
- Additions to the UI for quality-of-life improvements, e.g., features to save, change layer colours on the fly, select files, return to the desktop, etc.

These improvements, as well as a future user study to assess the acceptance of such novel concepts in 3D segmentation, will be the focus of future work.

References

Fiorillo, F., Jiménez Fernández-Palacios, B., Remondino, F., & Barba, S., 2013. 3D Surveying and modelling of the Archaeological Area of Paestum, Italy. *Virtual Archaeology Review*, 4, 55–60.

Fol, C. R., Murtiyoso, A., & Griess, V. C., 2022. Feasibility study of using virtual reality for interactive and immersive semantic segmentation of single tree stems. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLVIII-2/W1-2022, 95–99.

Fol, C. R., Shi, N., Overney, N., Murtiyoso, A., Remondino, F., & Griess, V. C., 2024. 3D Dataset Generation Using Virtual Reality for Forest Biodiversity. *International Journal of Digital Earth*, 17(1).

Franzluebbers, A., Li, C., Paterson, A., & Johnsen, K., 2022. Virtual Reality Point Cloud Annotation. Proceedings of the 2022 ACM Symposium on Spatial User Interaction.

Grilli, E., Farella, E. M., Torresani, A., & Remondino, F., 2019. Geometric features analysis for the classification of cultural heritage point clouds. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-2/W15, 541–548.

Grilli, E., & Remondino, F., 2019. Classification of 3D digital heritage. *Remote Sensing*, 11(7), 1–23.

Grilli, E., & Remondino, F., 2020. Machine Learning Generalisation across Different 3D Architectural Heritage. *ISPRS International Journal of Geo-Information*, 9(6), 379.

Kharroubi, A., Hajji, R., Billen, R., & Poux, F., 2019. Classification and Integration of Massive 3D Points Clouds in a Virtual Reality (VR) Environment. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-2/W17, 165–171.

Rehush, N., Abegg, M., Waser, L. T., & Brändli, U.-B., 2018. Identifying tree-related microhabitats in TLS point clouds using machine learning. *Remote Sensing*, 10(11), 1735.

Stathopoulou, E. K., Battisti, R., Cernea, D., Remondino, F., & Georgopoulos, A., 2021. Semantically derived geometric constraints for MVS reconstruction of textureless areas. *Remote Sensing*, 13(6), 1–19.

Suh, A., & Prophet, J., 2018. The state of immersive technology research: A literature analysis, *Computers in Human Behavior*, 86, 77-90.

Venn, L., & Mills, S., 2023. A VR Tool for Labelling 3D Data Sets. In W. Q. Yan, M. Nguyen, & M. Stommel (Eds.), *Image and Vision Computing*, pp. 262–271, Springer Nature Switzerland.

Yan, Z., Mazzacca, G., Rigon, S., Farella, E. M., Trybala, P., & Remondino, F., 2023. NeRFBK: A holistic dataset for benchmarking NeRF-based 3D reconstruction. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLVIII-1/W3-2023, 219–226.

Zhao, J., Wallgrün, J. O., LaFemina, P. C., Normandeau, J., & Klippel, A., 2019. Harnessing the power of immersive virtual reality - visualization and analysis of 3D earth science data sets. *Geo-Spatial Information Science*, 22(4), 237–250.

Appendix

The datasets used in this paper are all part of open data, which may be accessed for download as of 14 November 2024. The following data sources were used in this study:

- Flora: https://doi.org/10.3929/ethz-b-000694978
- Paestum: https://github.com/3DOM-FBK/NeRFBK
- Doss Trento: https://github.com/3DOM-FBK/NeRFBK
- Portici: https://github.com/3DOM-FBK/3DOM-Semantic-Facade

The VR labelling application used, Labelling Flora, is available in the following link: https://doi.org/10.5281/zenodo.13933004 (last accessed 14 November 2024). The RF4PCC method is available at https://github.com/3DOM-FBK/RF4PCC (last accessed 14 November 2024).