Transfer Learning Strategies for the Enhancement of Point Cloud Semantic Segmentation Processes in Landscape Heritage Contexts

Transfer Learning Strategies for the Enhancement of Point Cloud Semantic Segmentation Processes in Landscape Heritage Contexts

Marco Cappellazzo[1], Giacomo Patrucco[1*], Antonia Spanò[1,2]

[1] LabG4CH - Laboratory of Geomatics for Cultural Heritage. Department of Architecture and Design (DAD), Torino, Italy - (marco.cappellazzo, giacomo.patrucco, antonia.spano)@polito.it
[2] Polito FULL | The Future Urban Legacy Lab - OGR Tech, Torino, Italy

Keywords: Airborne LiDAR, Deep Learning, Semantic Segmentation, Landscape Heritage, Urban Legacies, Fine-tuning, Transfer Learning.

Abstract

In recent years, the increasing complexity of spatial data related to landscape heritage and urban legacies has led to a growing focus in research on enhancing the efficiency of management processes while simultaneously increasing their level of automation. In this context, a highly relevant solution is the use of so-called artificial intelligence, and more specifically, predictive models generated through Deep Learning techniques. However, a commonly observed critical issue concerns the limited generalisation capability of these models, which often fail to accurately recognise features in datasets that differ from those used during the training phase. In the present contribution, starting from a predictive model trained on airborne LiDAR point clouds belonging to a regional dataset (Sardinia, Italy), a transfer learning approach is proposed using new data (derived from the ISPRS benchmark dataset for semantic segmentation of Hessigheim – H3D) to improve the generalisation capabilities of the model. In order to assess the suitability of the proposed transfer learning strategy, a comparison between the classification performed by the original predictive model and the fine-tuned predictive model has been performed. Furthermore, the evaluation metrics have been calculated, evaluated, and discussed to quantitatively assess the improvement in terms of results, performance, and absolute gain between the different models tested. The proposed workflow supports scalable landscape and urban heritage monitoring by reducing human intervention in data management workflows while maintaining semantic consistency in available airborne laser scanner (ALS) data processing. While the contribution presents class limitations due to the constraints of the training data used for the first predictive model, the study demonstrates how transfer learning strategies can enhance the performance of semantic segmentation models when handling existing sparse data. This aligns well with scientific community efforts toward automated, efficient, and scalable heritage monitoring and documentation using remote sensing techniques.

1. Introduction

In today's society, the protection and conservation of cultural heritage (CH) has reached a level of complexity that is primarily conceptual, as it involves a wide range of different and disparate contexts and issues.
Since the post-war period of the last century, the recognition of CH as a key element of identity and collective memory, both local and global, has become more stable and substantial. This has been accompanied by concepts as the interdependence between tangible and intangible heritage (Bouhenaki, 2003), a growing and more mature awareness of anthropic and natural risk (U.N. Disaster Risk Reduction; Sendai Framework 2015, COPERNICUS), the promotion of sustainable tourism (UNESCO 2017), and the recognition of heritage as an economic driver for societies (UN, SDG11). These developments represent some of the most recent statements and concepts underlying the role of CH protection today.
Landscape heritage, or historical landscape, displays the strongest and most significant connections within the territory, as the relationship between human settlements and their placement on the territory follows recurring and recognisable patterns (Chapman, 2006).
In this sense, remote sensing (RS) methods, in their extraordinarily varied forms, have provided a significant non-invasive contribution to the documentation of landscape heritage (Forte & Campana, 2016). Following technological innovations, these methods have been able to offer different distance ranges of aerial and spaceborne acquisitions, using both active and passive sensors based on imaging and distance measurements, and operate across various bands of the electromagnetic spectrum, enabling analyses in both the visible and multi/hyperspectral domains.
The monitoring and protection of cultural heritage landscapes, including historic urban and landscape heritage sites, are essential for safeguarding cultural identity and ensuring sustainable territorial management. Heritage documentation often requires scalable, repeatable, and cost-effective methods to detect changes such as vegetation encroachment either on archaeological sites or urban centers, illegal land use transformations, and erosion processes threatening archaeological landscapes. In this context, airborne laser scanner (ALS) offers a practical tool for large-scale 3D documentation, but the effective use of these data in heritage management requires automated, robust semantic segmentation to minimize manual post-processing and enhance data structuring within institutional available data workflows.
In particular, the presented research focuses on ALS point clouds and the subsequent Digital Height Models (DHMs), which have played a decisive role in cartographic production and various fields of environmental studies investigating issues related to spatial and radiometric resolution, echo return information, density and, obviously, the possibility of boosting automation in analysis processes for deriving structured information (Mallet & Bretar, 2009).
The use of aerial platforms combined with full-waveform LiDAR, along with the use of Global Navigation Satellite Systems (GNSS) and Inertial Measurement Units (IMUs) (Shan & Toth, 2018) for the reconstruction of sensor trajectories and consequent georeferencing, has significantly improved the

https://doi.org/10.5194/isprs-annals-X-M-2-2025-47-2025

productivity of 3D data acquisition. This has enabled highly accurate 3D mapping for a wide range of purposes, for urban modeling (Nys et al., 2020), but also for microtopographic analyses that are particularly significant for mapping landscape heritage.

Among the critical issues arising from this development in the field of geo-information, particularly concerning multi-scale and multi-sensor approaches, one of the most significant challenges is undoubtedly the management of big spatial data. This issue becomes even more complex when considered from a multi-temporal collection perspective that has developed in parallel, posing significant challenges for common Geo-ICT infrastructures, which must manage these data with increasing effectiveness (Van Oosterom et al. 2015).

On the other hand, automated and integrated data structuring approaches aimed at efficiently deriving Digital Terrain Models (DTMs) and pursuing the automatic segmentation and classification of 3D data represent another crucial study scenario, as demonstrated by the significant development of the related literature (Chen et al., 2017).

Additionally, managing big spatial data is pivotal for applications ranging from city planning to environmental monitoring to autonomous navigation (Li et al., 2016). However, the volume of data, as well as the complexity of the information, requires effective data labelling automation, which is a key step towards developing AI-based analytics systems. In particular, semantic segmentation models for 3D point clouds have become essential (Matrone et al., 2020), by using neural networks (NNs) such as PointNet/PointNet++ (Qi et al., 2017a, 2017b), RandLA-Net (Hu et al., 2019), or recent transformer-based models like Point Transformer (Robert et al., 2023). Although they have great potential, deep learning models (DLMs) are subject to overfitting and data uncertainties, particularly driven by the scarcity of reference data (Weiss et al., 2016). In this sense, transfer learning pipelines are considered extremely crucial since the advent of DL, because they allow avoiding training models on new data from scratch, minimizing the human operator time in data labelling, while still achieving great performances (Pan & Yang, 2010; Sohail et al., 2025). In fact, in order to augment performance and generalisation capabilities, it is possible to first train DLMs on larger, related datasets and fine-tune them on specific target domains, making it possible to achieve higher performance with fewer annotated examples (Iman et al., 2023).

Finally, aligning with the objectives of the CIPA community (CIPA Heritage Documentation), this work explores the potential of DLMs and transfer learning strategies to automate and optimize the processing of ALS point clouds for landscape and urban heritage monitoring.

1.1 Previous research experience and present research aims

The present contribution falls within this last research sector, starting from an extensive previous research experience developed by the authors, (Cappellazzo et al., 2024; Cappellazzo, 2025) , that aimed to automate data structuring by leveraging 3D ALS point clouds available from regional spatial data infrastructures by applying machine learning (ML) techniques to enhance the detection of Sardinia’s coastal landscape heritage, which is characterised both by isolated wild landscapes (Figure 1) and by urban legacies. Documenting and preserving isolated coastal landscapes and urban legacies is crucial for maintaining cultural identity, fostering community belonging, and promoting sustainable development. In this sense, throughout participatory mapping, documentation, and data structuring tools, cities can safeguard their layered memories while continuously adapting to contemporary needs. The present contribution methodological investigation is based on the analysis and study of foundational knowledge within the heritage conservation disciplines. Specifically, the present research has examined the multidisciplinary approaches developed by research groups of the University of Cagliari (Giannattasio et al., 2020; Fiorino et al., 2021). The focus of this research is to document and investigate Sardinian construction techniques, since architectural elements within the historic Sardinian landscape are notably diverse, both chronologically and in terms of typology and construction methods. Starting from this premise, the restoration research group from Cagliari has developed a series of in-depth studies focused on the knowledge and recognition of the local built heritage, addressing interdisciplinary methodologies aimed at chronological and typological analysis to establish effective tools for the conservation of this heritage. The categories of assets investigated are grouped into four macro-categories, starting with military defensive architecture, religious architecture, civil and vernacular architecture, and historical centers (Giannattasio et al., 2020). In this scenario, the topic of historical defensive architectures and military landscapes of Sardinia (Fiorino et al., 2021) has been specifically addressed by issuing a shared methodology. In this sense, the University of Cagliari's research requires exploring the geospatial relations between military architecture with its construction techniques, the actual and historic geographic and sociopolitical context, and the morpho-typological characteristics, by structuring a geographic database. This has been made necessary because of the slow and continuous process of the defensive infrastructure development of the island, driven by the evolving needs of the defensive functions over the centuries.

Figure 1. Sevo Tower. Cabras, Italy (cr. fortificazioni.net).

For these reasons, the previous research from the authors explored the development of a predictive deep learning (DL) method for the automatic segmentation of DTMs, aimed at classifying the coastal military heritage. However, this method required further fine-tuning and enhancement to apply to the landscape heritage's scope.

Certainly, many landscape heritage sites, both in Sardinia and elsewhere, can be extensively studied, and in many cases, their history, transformation, and cultural values are known and well-documented. The final goal of this kind of study is not to generate small-scale models of specific samples of widespread or scattered landscape heritage. This is especially true considering that 3D models, acquired through terrestrial methods, can provide denser and more accurate documentation products. Rather, the research aims to develop the potential for automatic and rapid archiving at the scale of regional cartography, aiming to improve the integration of public and widely distributed databases and provide public administrations with sustainable and innovative tools for comprehensive territorial and landscape management. In the specific case of the presented research, the goal is to propose a method to enhance the effectiveness of these processes.

Starting from the predictive model trained during the aforementioned research activities (Cappellazzo et al. 2024; Cappellazzo 2025), a transfer learning approach was tested using new data in order to evaluate how such a strategy could improve the adaptability of predictive models while also increasing their flexibility and efficiency.

2. Data and methodology

2.1 Primary data

The present contribution is thus focused on transfer learning pipelines starting from a previously trained semantic segmentation model, DLM1 (Cappellazzo et al. 2024).
The DLM1 model was trained using RandLA-Net (Hu et al., 2019) architecture and validated on an airborne LiDAR dataset pertaining to the area of Cagliari (2 pts/m²) and tested on another dataset representing Alghero area (10 pts/m²). These datasets belong to a regional dataset (Sardinia Region) derived from two different airborne LiDAR surveys carried out in 2008, while Cagliari data were acquired with an ALTM Gemini sensor at a 1400 m average above-ground level (AGL), Alghero data acquisition involved a Riegl LMS-Q560 full-waveform sensor, operating at a 500 m AGL. Given that the datasets have been characterized by low densities, data uncertainties, and a lack of related reflectance information, the preparation of training data has involved the development of a satellite imagery data-fusion approach combined with an unsupervised filtering algorithm.
As described in (Cappellazzo et al. 2024), a sample of the semi-automatic annotated data (≈ 2.7 km^2) has been used for the training of a DLM, splitting the sample into 78% for the training and 22% for validation, applying a five-class semantic segmentation scheme (never classified; ground; high vegetation; building; water). Validation of the model achieved a macro-average of 0.95 accuracy, 0.80 precision, 0.73 recall, and 0.74 F1-score; however, slight overfitting, even if the results across four independent test areas reached 0.96 accuracy, 0.89 precision, 0.77 recall, and 0.80 F1-score.
Validation of the model achieved a macro-average of 0.95 accuracy, 0.80 precision, 0.73 recall, and 0.74 F1-score; however, slight overfitting, even if the results across four independent test areas reached 0.96 accuracy, 0.89 precision, 0.77 recall, and 0.80 F1-score.
The model demonstrated strong transferability across different acquisition densities and territorial morphologies. However, this has been proved to be effective in urban areas with similar morphologies, and the model failed to demonstrate generalisation capabilities in extremely different contexts.
In fact, in the present contribution, the ISPRS benchmark dataset for semantic segmentation of Hessigheim (H3D) (Kölle et al., 2018) was used to specifically "stress-test" the DLM1 on different data. H3D benchmark consists of high-density LiDAR data (800 pts/m^2) originating from UAS-based LiDAR and digital photogrammetry. Due to the high density of the point clouds, along with high-resolution RGB data (2-3 cm GSD), it is possible to estimate high-granularity details in the urban scene. The specific aim of H3D is to provide properly labelled data sets, useful to test and improve point cloud semantic segmentation methods for geospatial applications. Moreover, one of the aims of the WG II/2 was also to deal with data semantics, providing a class scheme coherent with the resolution of data.
A significant difference between the two datasets is that the regional dataset from the Sardinia Region includes points labelled as belonging to the "water" class during the annotation phase preceding the training, whereas this class is absent in the Hessigheim dataset. Consequently, during the experimentation carried out in the present paper, the "water" class was excluded and not considered.
Still, as shown in Figure 2, where two aerial images of Cagliari (where the first predictive model was trained) and Hessigheim are shown, the two areas exhibit substantial differences in morphology, urban fabric, and territorial context. Within this framework, as further discussed in Section 2.2, the DLM1 model encountered features that are significantly different from those encountered during its original training phase.

a) Cagliari, Italy (cr. Wikimedia)

b) Hessigheim, Germany (cr. Wikimedia)

Figure 2. Comparison between the Cagliari city centre (a) and the Hessigheim (b) settlements' morphologies. From the aerial photographs, it is possible to observe the important differences between the two urban configurations in density and construction typologies. This difference reflects the available digital data.

The H3D dataset thus served as a complementary support for the verification of the original DLM1 performances, aligning with the objectives of scalable heritage documentation. The Sardinia ALS dataset from which the original backbone model has been trained represents the typical situation of regional available datasets, characterised by low-density airborne data and limited possibilities for class annotations, reflecting the constraints over data resolution. Yet, the ISPRS H3D benchmark significantly differs in terms of urban morphology and point density and is necessary for issuing a stress-test of the generalisation capacity of the predictive model and evaluating transfer learning strategies under different resolution conditions. The complementarity of the two datasets is an essential condition for a robust assessment of model adaptability, while acknowledging at the same time the differences in dataset characteristics and their implications for urban and landscape scale heritage data processing workflows.

2.2 Class remapping and subsampling strategies

In order to carry out the fine-tuning process using the H3D benchmark dataset, starting from the predictive model trained on the Sardinian case study, the class labels were remapped. This step is necessary because the fine-tuning data must be labelled according to the class scheme of the predictive model, allowing the model to learn effectively from the new input data.

https://doi.org/10.5194/isprs-annals-X-M-2-2025-47-2025

Table 1 shows the class remapping between the Hessigheim 3D data labels and the scheme used to annotate the original semantic segmentation model, as well as the corresponding class descriptions.

Original class	Description	Target class	Description
C00	Low Vegetation	C02	Ground
C01	Impervious Surface	C02	Ground
C02	Vehicle	C00	Never Classified
C03	Urban Furniture	C00	Never Classified
C04	Roof	C06	Building
C05	Façade	C06	Building
C06	Shrub	C00	Never Classified
C07	Tree	C05	High Vegetation
C08	Soil/Gravel	C02	Ground
C09	Vertical Surface	C00	Never Classified

Table 1. Hessigheim 3D data class schema remapping.

The point clouds from the Sardinian dataset, used to train the original predictive model, were characterised by a density of 2 pts/m^2. However, to evaluate the effectiveness of the fine-tuning strategy and the adaptability of the resulting model, multiple training experiments were carried out using data from the H3D dataset subsampled at varying densities.
Specifically, the first subsampling was performed to achieve the same density of the Sardinian point clouds; furthermore, two additional subsampled datasets were considered with a point spacing of 0.5 m (resulting in a point cloud with an average density of 10 pts/m^2) and 0.15 m (resulting in a point cloud with an average density of 50 pts/m^2).
The principal characteristics of this subsampling process (point spacing and average density) are reported in Table 2.

SS instance	Point spacing [m]	Average density
SS1	no subsampling	≈ 800 pts/m^2
SS2	1	≈ 2 pts/m^2
SS3	0.5	≈ 10 pts/m^2
SS4	0.15	≈ 50 pts/m^2

Table 2. Hessigheim 3D data subsampling strategies (SS).

The choice made by the authors of subsampling data reflects practical and real-world situations. In fact, low resolutions (2-10 pts/m^2, lack of data, and data uncertainties are often typical of available regional remote sensing airborne datasets used by administrations for mapping and territorial monitoring. While the previous study (Cappellazzo, 2024) was focused on exploiting regional available data also for landscape-scale heritage monitoring, the present contribution also focuses on higher resolutions (>50 pts/m²) to simulate tailored dense acquisitions that can be employed for specific monitoring of sensitive heritage areas requiring more detailed analysis. In this sense, this model could be considered a preliminary model for point cloud semantic segmentation, which could prepare raw data for a more detailed manual annotation, and consequently for the training of a tailored model.

2.3 Classification tests with the original DLM

The predictive model previously trained on data from Sardinia was applied to classify the point clouds belonging to the H3D dataset and assess its generalisation capability. The test was conducted on the original resolution data (SS1) and subsampled point clouds characterised by different densities (SS2, SS3, SS4), as visible from Figure 3.
From the visual comparison, it is clear how the model failed to achieve a correct classification of the input data, empirically demonstrating the supposed overfitting phenomenon observed during the training process. In fact, although the good results were obtained with an external Sardinian test dataset, the urban conformation of the two case studies presents important differences, as also shown in Figure 2.

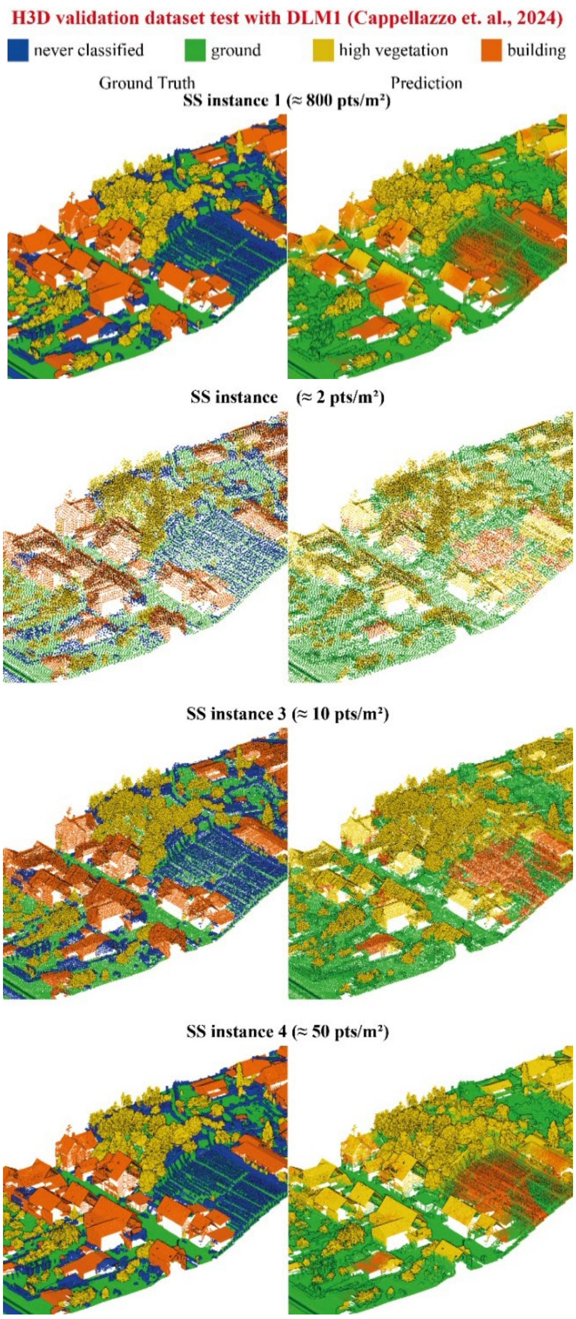

Figure 3. Ground truth/predictions comparison of the semantic segmentation model trained on the Sardinia dataset, tested on the validation set of the Hessigheim benchmark (H3D) (Kölle et al., 2021).

From the visual comparison, it is clear how the model failed to achieve a correct classification of the input data, empirically demonstrating the supposed overfitting phenomenon observed during the training process. In fact, although the good results were obtained with an external Sardinian test dataset, the urban conformation of the two case studies presents important differences, as also shown in Figure 2.

The evaluation metrics – calculated from the true/false positives/negatives ratio reported in the Confusion Matrix – are reported in Table 3 (classification of SS1 dataset), Table 4 (classification of SS2 dataset), Table 5 (classification of SS3 dataset), and Table 6 (classification of SS4 dataset).

DLM1 on Validation SS1	Accuracy	Precision	Recall	F1
Never Classified	0.90	0.24	0.00	0.00
Ground	0.83	0.77	0.90	0.83
High vegetation	0.84	0.38	0.81	0.52
Building	0.78	0.71	0.49	0.58
Macro Average	0.83	0.53	0.55	0.48

Table 3. Evaluation metrics achieved on the validation dataset using the predictive model trained from the Sardinia point cloud (full resolution data).

DLM1 on Validation SS2	Accuracy	Precision	Recall	F1
Never Classified	0.83	0.27	0.00	0.00
Ground	0.78	0.61	0.88	0.72
High vegetation	0.67	0.41	0.86	0.56
Building	0.72	0.34	0.04	0.08
Macro Average	0.75	0.41	0.45	0.34

Table 4. Evaluation metrics achieved on the validation dataset using the predictive model trained from the Sardinia point cloud (subsampled data – 2 pts/m^2).

DLM1 on Validation SS3	Accuracy	Precision	Recall	F1
Never Classified	0.82	0.00	0.00	0.00
Ground	0.77	0.59	0.86	0.70
High vegetation	0.69	0.47	0.86	0.60
Building	0.73	0.32	0.06	0.11
Macro Average	0.76	0.34	0.45	0.35

Table 5. Evaluation metrics achieved on the validation dataset using the predictive model trained from the Sardinia point cloud (subsampled data – 10 pts/m^2).

DLM1 on Validation SS4	Accuracy	Precision	Recall	F1
Never Classified	0.82	0.33	0.00	0.00
Ground	0.77	0.59	0.85	0.70
High vegetation	0.71	0.49	0.87	0.63
Building	0.76	0.38	0.09	0.15
Macro Average	0.76	0.45	0.45	0.37

Table 6. Evaluation metrics achieved on the validation dataset using the predictive model trained from the Sardinia point cloud (subsampled data – 50 pts/m^2).

As expected, the results obtained from this preliminary classification highlight a low performance, indicating a limited capability of the predictive model to identify the classes of the analysed data correctly. The class least affected by the differences between the training data and the classified data is the ground class, across all four tests performed. This is the only class that consistently shows sufficient values regarding Precision, Recall, and, consequently, F1-score.
On the other hand, the class that shows the highest accuracy in all four cases is the Never Classified class, but this is due to the imbalance between the high value of True Negatives and the very low number of correctly identified True Positives. This aspect is clearly highlighted by the extremely low values of Precision and Recall.
The second class (after Ground) to show the best performance is High Vegetation, despite the extremely low Precision value across all four cases. This characteristic highlights a tendency of the model to fail to correctly identify many of the positive elements belonging to this class, resulting in a high number of false negatives. In contrast, the extremely low Recall value for the Building class underlines how this specific class is generally underpredicted.
It is no coincidence that for the Building class, results are characterised by very low metrics, given that the urban fabric of the two areas considered is extremely different, as evidenced by the various building types and the varying roof pitch angles.
In this case, it is important to underline that the data to be classified significantly differs from the training data. In this scenario, the used predictive model does not have sufficient generalisation capability to make accurate predictions.

2.4 Transfer learning training tests

Four additional fine-tuning training workflows were developed to further evaluate the robustness and adaptability of the semantic segmentation predictive models, leveraging the DLM1 as a backbone model. The transfer learning test experiments were conducted by adopting the four different subsampling strategies described in Table 1, applied to the input H3D point clouds.
The training experiments are resumed in Table 7.

TL instance	Class Remapping	Sub Sampling	Block size [m]
TL1	Yes	No	100
TL2	Yes	SS1	100
TL3	Yes	SS2	100
TL4	Yes	SS3	100

Table 7. Transfer learning (TL) test strategies.

Each training process maintained the same NN of the original model (RandLA-Net), and training data block extension, while modifying only the sampling stage prior to feature extraction. Specifically, the input point clouds were independently processed using a distance-based uniform sampling. In each case, the same class schema and some common hyperparameters (batch size of 10 blocks, a starting learning rate of 0.001, and a cross-entropy loss function) were adopted. The learning rate optimiser was employed throughout, and training was conducted over a maximum of 50 epochs for each model with early stopping criteria based on F1 and best validation loss. In this sense, block dimensions and training data preparation are strongly influenced by the backbone model, while hyperparameters such as learning rate, batch size, and model selection criteria are lessons learned from the study of the state-of-the-art previous research experiences (Cappellazzo, 2024).
The choice to explore four different subsampling strategies was mainly driven by the need to understand not just how they affect overall segmentation accuracy, but also how sensitive the models

https://doi.org/10.5194/isprs-annals-X-M-2-2025-47-2025 | © Author(s) 2025. CC BY 4.0 License.

are to small variations in data regarding the point clouds topology, the geometric details, and finer structures within the point clouds. The distance-based sampling offered a straightforward test of how imposing a regular structure might impact model performance, even though it risks losing finer detail. All training experiments were run on an NVIDIA RTX 3090 GPU, which provided enough memory to handle large spatial tiles efficiently without forcing memory consumption.

3. Results and discussion

At the end of the fine-tuning procedures described in the previous section, the Validation Dataset for TL1, TL2, TL3, and TL4 was classified using the fine-tuned predictive models in order to generate the Confusion Matrix, with the aim of assessing the effectiveness of the newly trained model. Consequently, the evaluation metrics (Accuracy, Precision, Recall, and F1) were calculated. The evaluation metrics related to each training are reported in Tables 8, 9, 10, and 11. In this case, it is evident that the order of magnitude of the performance exhibited by the models after fine-tuning is significantly superior to that of the previous model.

TL1 on Validation SS1	Accuracy	Precision	Recall	F1
Never Classified	0.91	0.56	0.44	0.49
Ground	0.91	0.86	0.97	0.91
High vegetation	0.96	0.83	0.78	0.80
Building	0.92	0.91	0.82	0.86
Macro Average	0.92	0.79	0.75	0.77

Table 8. Evaluation metrics achieved on the validation dataset (fine-tuning, full-resolution data).

TL2 on Validation SS2	Accuracy	Precision	Recall	F1
Never Classified	0.84	0.54	0.45	0.49
Ground	0.90	0.86	0.82	0.84
High vegetation	0.81	0.57	0.94	0.71
Building	0.85	0.90	0.50	0.64
Macro Average	0.85	0.72	0.68	0.67

Table 9. Evaluation metrics achieved on the validation dataset (fine tuning, subsampled data – 10 pts/m^2).

TL3 on Validation SS3	Accuracy	Precision	Recall	F1
Never Classified	0.88	0.73	0.51	0.60
Ground	0.94	0.90	0.92	0.91
High vegetation	0.93	0.84	0.94	0.89
Building	0.93	0.84	0.88	0.86
Macro Average	0.92	0.83	0.81	0.82

Table 10. Evaluation metrics achieved on the validation dataset (fine tuning, subsampled data – 2 pts/m^2).

TL4 on Validation SS4	Accuracy	Precision	Recall	F1
Never Classified	0.88	0.66	0.66	0.66
Ground	0.94	0.89	0.93	0.91
High vegetation	0.94	0.89	0.90	0.90
Building	0.95	0.92	0.84	0.88
Macro Average	0.93	0.84	0.83	0.83

Table 11. Evaluation metrics achieved on the validation dataset (fine tuning, subsampled data – 50 pts/m^2).

The results obtained from the classification using the original predictive model were compared with those achieved by the model fine-tuned on the H3D dataset. In particular, Tables 12, 13, 14 and 15 report the percentage differences in evaluation metrics, highlighting a significant performance improvement for

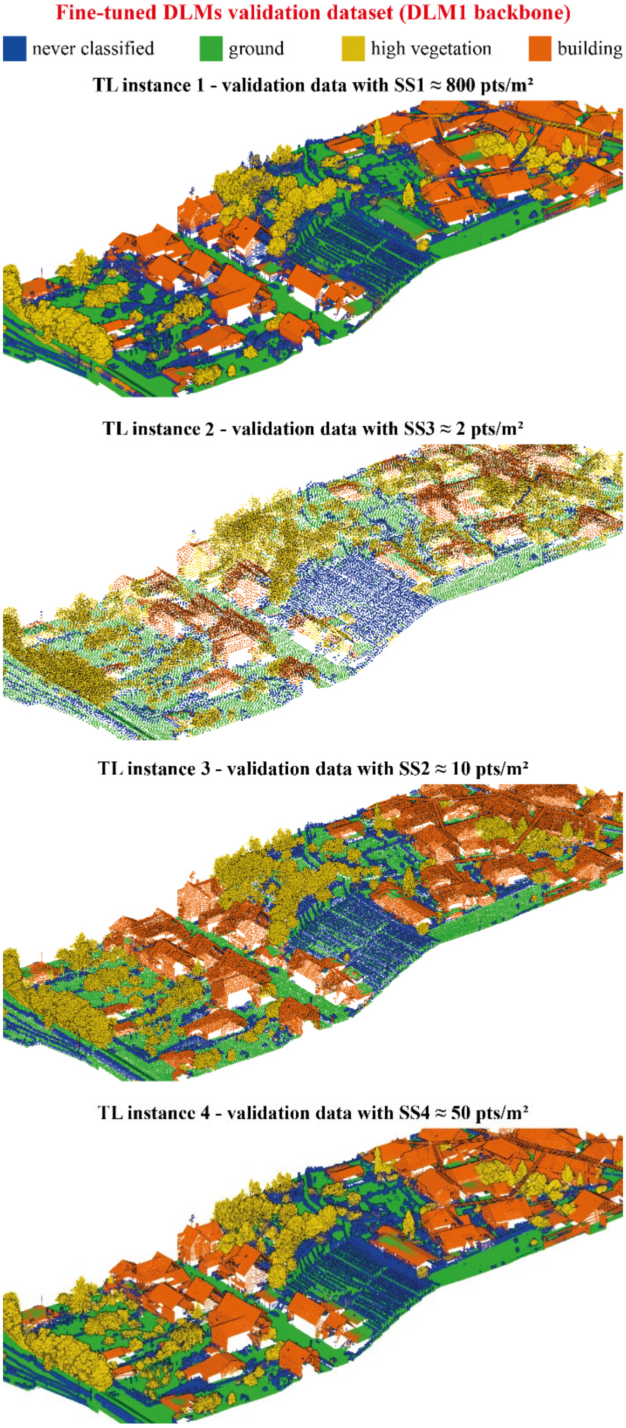

Figure 4. Prediction results of the semantic segmentation models were trained using the DLM1 as the backbone, tested on the validation dataset of H3D (Kölle et al., 2021).

the second model, which learned the dataset-specific features from the fine-tuning data in order to assign each element to its corresponding class correctly. Such performance was clearly unachievable with the previous model, due to the significant morphological and typological differences between the elements of the H3D dataset and those used for training the Sardinian model, caused by the extremely different characteristics of the two urban fabrics considered.

TL1 absolute gain	Evaluation metric percentage differences (original model/fine-tuned model)			
	Accuracy	Precision	Recall	F1
Never Classified	+1%	+32%	+44%	+49%
Ground	+8%	+9%	+7%	+8%
High vegetation	+12%	+45%	-3%	+29%
Building	+14%	+20%	+33%	+28%
Macro Average	+9%	+26%	+20%	+29%

Table 12. Comparison between the evaluation metrics achieved from the original model and the fine-tuned model, evidencing a significant improvement in terms of performance (TL1).

TL2 absolute gain	Evaluation metric percentage differences (original model/fine-tuned model)			
	Accuracy	Precision	Recall	F1
Never Classified	+1%	+26%	+45%	+49%
Ground	+12%	+25%	-6%	+12%
High vegetation	+14%	+15%	+8%	+15%
Building	+13%	+56%	+46%	+57%
Macro Average	+10%	+31%	+23%	+33%

Table 13. Comparison between the evaluation metrics achieved from the original model and the fine-tuned model, evidencing a significant improvement in terms of performance (TL2).

TL3 absolute gain	Evaluation metric percentage differences (original model/fine-tuned model)			
	Accuracy	Precision	Recall	F1
Never Classified	+6%	+73%	+51%	+60%
Ground	+17%	+31%	+6%	+21%
High vegetation	+24%	+37%	+8%	+28%
Building	+20%	+53%	+82%	+76%
Macro Average	+17%	+49%	+37%	+46%

Table 14. Comparison between the evaluation metrics achieved from the original model and the fine-tuned model, evidencing a significant improvement in terms of performance (TL3).

TL4 absolute gain	Evaluation metric percentage differences (original model/fine-tuned model)			
	Accuracy	Precision	Recall	F1
Never Classified	+6%	+33%	+66%	+66%
Ground	+17%	+30%	+8%	+21%
High vegetation	+23%	+40%	+3%	+27%
Building	+19%	+54%	+75%	+73%
Macro Average	+17%	+39%	+38%	+46%

Table 15. Comparison between the evaluation metrics achieved from the original model and the fine-tuned model, evidencing a significant improvement in terms of performance (TL4).

From the analysis of the data presented above, it is clear that the most significant performance improvements are observed for the Building class, which stands out as the class that has benefited the most from this transfer learning strategy, with increases in terms of performance of up to +56% for Precision (TL2), up to 82% for Recall (TL3), and up to +76% for F1-score (TL3).
Overall, all classes show satisfactory metrics, with the exception, as expected, of the Never Classified class. However, it still shows an increase, compared to the classification performed with the previous predictive model, of several percentage points in performance, especially in terms of Recall and F1-score.
One aspect that emerges when looking at the Macro Average of the calculated metrics is that, despite the original predictive model being trained with point clouds characterised by a density of approximately 2 pts/m^2, and therefore the dataset being the closest to the original one, the most positive performance is observed on point clouds with higher density. In particular, considering the classes Ground, High Vegetation and Building, the dataset with the overall highest metrics are TL3 and TL4 (characterised by a density of approximately 10 pts/m^2 and 50 pts/m^2 respectively).
This may be attributed to the higher resolution of the data and, consequently, to the improved spatial representation of complex elements such as buildings, structures, or landscape features. In fact, denser point clouds have provided a more effective input for the new predictive model that has been generated according to the strategies outlined in Section 2.4. In conclusion, the improvements in evaluation metrics, achieved after the fine-tuning, demonstrate that the method can be applied for scalable heritage monitoring. In fact, since most of the models achieved an F1-score > 0.80 with an absolute gain in the range of +30%/+40% on the building class, the models can thus detect changes in historic urban fabric with an improved generalisation capability, reducing the need for manual post-processing editing by human operators. Similarly, the metrics for the vegetation class demonstrate the models being able to support the monitoring of high-vegetation invasion on archaeological and landscape sites, facilitating heritage sites management and conservation.

4. Conclusions

In the scenario of semantic segmentation of point clouds, and more specifically considering point clouds used for documenting landscape heritage and urban legacy, which represents the final goal for the generation of the Sardinian predictive model as stated in (Cappellazzo et al. 2024), the potential of fine-tuning strategies for transfer learning purposes offers significant opportunities. In fact, the present contribution demonstrated that by systematically varying the point cloud resolution within the training data preparation stage while keeping the network architecture and optimisation parameters constant, it is possible to understand the influence of subsampling strategies based on the performance and generalisation capabilities of each model. The research was aiming to develop a controlled, modular methodology aimed at optimising LiDAR data structuring workflows by employing transfer learning pipelines. In this sense, the absolute gain in performance metrics shown in Tables 12-15 demonstrates the reliability of fine-tuning to improve the generalisation capabilities of DLMs.

References

Bouhenaki, M. 2003. The interdependency of the tangible and intangible cultural heritage. 14th ICOMOS General Assembly and International Symposium: 'Place, Memory, Meaning: Preserving Intangible Values in Monuments and Sites', 27 – 31 Oct 2003, Victoria Falls, Zimbabwe. [Conference or Workshop Item]. https://openarchive.icomos.org/id/eprint/468/

Cappellazzo, M. 2025. Semantic classification of airborne LiDAR data. Built and landscape heritage analysis in 2D/3D GIS environment [PhD thesis, Politecnico di Torino]. https://iris.polito.it/handle/11583/2998057

Cappellazzo, M., Patrucco, Spanò, A., 2024: ML Approaches for the Study of Significant Heritage Contexts: An Application on Coastal Landscapes in Sardinia. *Heritage*, 7(10), 5521-5546. doi.org/10.3390/heritage7100261.

Chapman, H., 2006: Landscape archaeology and GIS. The History Press, Briscombe Port, United Kingdom.

Chen, Z., Gao, B., Devereux, B., 2017: State-of-the-Art: DTM Generation Using Airborne LIDAR Data. *Sensors*, 17(1), 150. doi.org/10.3390/S17010150

CIPA Heritage Documentation (2025), *What is CIPA?* https://www.cipaheritagedocumentation.org/about/whatiscipa/ [Accessed: 11/07/2025]

Copernicus Climate Change Service (C3S). https://www.copernicus.eu/en/news/news/observer-copernicus-preservation-global-cultural-heritage-sites
Accessed April 2025.

Fiorino, D. R. (2021). Sinergies. Interinstitutional experiences for the rehabilitation of military areas (Vol. 1). UNICApress.
Forte, M., & Campana, S. (Eds.). 2016. Digital Methods and Remote Sensing in Archaeology. Springer International Publishing. https://doi.org/10.1007/978-3-319-40658-9

Giannattasio, C., Bartolomucci, C., & Pretelli, M. (2020). Arte muraria tradizionale in Sardegna : conoscenza conservazione miglioramento. Gangemi.

Iman, M., Arabnia, H. R., & Rasheed, K. 2023. A Review of Deep Transfer Learning and Recent Advancements. Technologies, 112), 40. https://doi.org/10.3390/technologies11020040

Hu, Q., Yang, B., Xie, L., Rosa, S., Guo, Y., Wang, Z., Trigoni, N., & Markham, A. 2019. RandLA-Net: Efficient Semantic Segmentation of Large-Scale Point Clouds. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 11105–11114. https://doi.org/10.1109/CVPR42600.2020.01112

Kölle, M., Laupheimer, D., Schmohl, S., Haala, N., Rottensteiner, F., Wegner, J. D., & Ledoux, H. 2021. The Hessigheim 3D (H3D) benchmark on semantic segmentation of high-resolution 3D point clouds and textured meshes from UAV LiDAR and Multi-View-Stereo. ISPRS Open Journal of Photogrammetry and Remote Sensing, 1, 100001. https://doi.org/10.1016/j.ophoto.2021.100001

Li, S., Dragicevic, S., Castro, F. A., Sester, M., Winter, S., Coltekin, A., Pettit, C., Jiang, B., Haworth, J., Stein, A., & Cheng, T. 2016. Geospatial big data handling theory and methods: A review and research challenges. ISPRS Journal of Photogrammetry and Remote Sensing, 115, 119–133. https://doi.org/10.1016/J.ISPRSJPRS.2015.10.012

Mallet, C., Bretar, F., 2009: Full-waveform topographic LiDAR: State-of-the-art. *ISPRS Journal of Photogrammetry and Remote Sensing*, 64(1), 1–16. 10.1016/j.isprsjprs.2008.09.007.

Matrone, F., Grilli, E., Martini, M., Paolanti, M., Pierdicca, R., & Remondino, F. 2020. Comparing Machine and Deep Learning Methods for Large 3D Heritage Semantic Segmentation. ISPRS International Journal of Geo-Information 2020, Vol. 9, Page 535, 9(9), 535. https://doi.org/10.3390/IJGI9090535
Nys, G. A., Poux, F., & Billen, R. 2020. CityJSON Building Generation from Airborne LiDAR 3D Point Clouds. ISPRS International Journal of Geo-Information 2020, Vol. 9, Page 521, 9(9), 521. https://doi.org/10.3390/IJGI9090521

Pan, S. J., & Yang, Q. 2010. A survey on transfer learning. IEEE Transactions on Knowledge and Data Engineering, 22(10), 1345–1359. https://doi.org/10.1109/TKDE.2009.191

Patil, A., Malla, S., Gang, H., Chen, Y.-T. H. Gang and Y. -T. Chen, 2019: The H3D Dataset for Full-Surround 3D Multi-Object Detection and Tracking in Crowded Urban Scenes. *Proceedings of International Conference on Robotics and Automation (ICRA) 2019*, 9552-9557. doi.org/10.1109/ICRA.2019.8793925.

Qi, C. R., Su, H., Mo, K., & Guibas, L. J. 2017). PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation.

Qi, C. R., Yi, L., Su, H., & Guibas, L. J. 2017. PointNet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space. Advances in Neural Information Processing Systems, 2017-December, 5100–5109. https://arxiv.org/abs/1706.02413v1

Robert, D., Raguet, H., & Landrieu, L. 2023. Efficient 3D Semantic Segmentation with Superpoint Transformer. Proceedings of the IEEE International Conference on Computer Vision, 17149–17158. https://doi.org/10.1109/ICCV51070.2023.01577

Shan, J., Toth, C. K., 2018: Topographic Laser Ranging and Scanning. CRC Press, Boca Raton, Florida, USA. doi.org/10.1201/9781315154381.

Sohail, S. S., Himeur, Y., Kheddar, H., Amira, A., Fadli, F., Atalla, S., Copiaco, A., & Mansoor, W. 2025. Advancing 3D point cloud understanding through deep transfer learning: A comprehensive survey. Information Fusion, 113, 102601. https://doi.org/10.1016/J.INFFUS.2024.102601

Van Oosterom, P. J. M, Martinez-Rubi, O., Ivanova, M., Horhammer, M., Geringer, D., Ravada, S., Tijssen, T., Kodde, M., Gonçalves, R., 2015: Massive point cloud data management: Design, implementation and execution of a point cloud benchmark. *Computers & Graphics*, 49, 92-125. doi.org/10.1016/j.cag.2015.01.007

U. Department of Economic and Social Affairs, "Disaster risk reduction." Disaster risk reduction | Department of Economic and Social Affairs. Accessed April 2025.

UN, "SUSTAINABLE DEVELOPMENT GOAL 11 Make Cities and Human Settlements Inclusive, Safe, Resilient And Sustainable A GUIDE TO ASSIST NATIONAL AND LOCAL GOVERNMENTS TO MONITOR AND REPORT ON SDG GOAL 11+ INDICATORS," New York. https://sdgs.un.org/topics/disaster-risk-reduction#description Accessed April 2025.

UNESCO 2011. Recommendation on the Historic Urban Landscape. Accessed April 2025.

UNESCO, 2017. World Heritage and Sustainable Tourism Programme. Accessed April 2025.

United Nations Office for Disaster Risk Reduction. Sendai framework for disaster risk reduction 2015–2030. Geneva: United Nations Office for Disaster Risk Reduction; 2015. Accessed April 2025.

Weiss, K., Khoshgoftaar, T. M., & Wang, D. D. 2016. A survey of transfer learning. Journal of Big Data, 3(1), 1–40. https://doi.org/10.1186/S40537-016-0043-6/TABLES/6