IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 25 November 2025, accepted 11 December 2025, date of publication 22 December 2025, date of current version 2 January 2026.

Digital Object Identifier 10.1109/ACCESS.2025.3647154

APPLIED RESEARCH

Deep Learning for Ultra-Large-Scale Semantic Segmentation of Geographic 3D Point Clouds With Missing Labels

ALBERTO M. ESMORÍS[1,2], **MIGUEL YERMO**[3,4], **SILVIA R. ALCARAZ**[3], **(Member, IEEE), SAMUEL SOUTULLO**[3], **AND FRANCISCO F. RIVERA**[3,4]

[1]Departamento de Matemáticas and CITMAga, Universidade de Santiago de Compostela, 15782 Santiago de Compostela, Spain
[2]3DGeo Research Group, Institute of Geography, Heidelberg University, 69120 Baden-Württemberg, Germany
[3]Centro Singular de Investigación en Tecnoloxías Intelixentes, Universidade de Santiago de Compostela, 15782 Santiago de Compostela, Spain
[4]Departamento de Electrónica e Computación, Universidade de Santiago de Compostela, 15782 Santiago de Compostela, Spain

Corresponding author: Alberto M. Esmorís (alberto.esmoris.pena@usc.es)

This work was supported in part by the Deutsche Forschungsgemeinschaft (DFG), German Research Foundation, by the Project VirtuaLearn3D under Grant 496418931; in part by the Agencia Estatal de Investigación (AEI), Spain, Ministerio de Ciencia, Innovación y Universidades (MCIN)/AEI/10.13039/501100011033 under Grant PID2019-104834GB-I00 and Grant PID2022-141623NB-I00; in part by the Xunta de Galicia-Consellería de Cultura, Educación, Formación Profesional e Universidades (Centro de investigación de Galicia Accreditation) under Grant 2024-2027 ED431G-2023/04 and Grant ED431C-2022/016; and in part by the European Union (European Regional Development Fund-ERDF/EU).

ABSTRACT Semantic segmentation of 3D point clouds is a critical task essential for research and industry in a wide variety of domains. Most works until now use datasets where the concept of large-scale varies between 1.17 and 9, 261 million points and 0.31 and 250 square kilometers. In this research, we push the limits of 3D deep learning by considering approximately 36, 369 million points and 29, 557 square kilometers in our case study. To the best of our knowledge, it is the largest geographic region ever classified with neural networks for 3D point clouds in scientific works. The main contributions of our research are: 1) The first published results on semantic segmentation of ultra-large-scale and low-resolution airborne laser scanning point clouds, 2) adapted neural networks to work under missing labels and high measurement error conditions, 3) the introduction of class ambiguity as a more robust uncertainty measurement compared to entropy, and 4) an open-source deep learning framework for 3D point clouds that enables the reproducibility of our experiments and further studies.

INDEX TERMS 3D point clouds, deep learning, large-scale, LiDAR, semantic segmentation, semi-supervised training.

I. INTRODUCTION

Three-dimensional point clouds are widely used for various applications like vegetation analysis for agricultural and forestry purposes, archaeology, city planning and maintenance, building monitoring, road surveys, topographic terrain analysis, bathymetry, robotics, and more. Typically, but not necessarily, 3D point clouds are acquired using Light Detection and Ranging (LiDAR) sensors. This technology obtains highly accurate 3D measurements through laser scanning campaigns [1].

Nowadays, one of the most relevant tasks in point cloud processing is point-wise classification. However, most scientific works use tiny datasets compared to the deep learning standards that demand a minimum of 5, 000 labeled examples per class and expect optimal performance with ten million examples per class [2]. These numbers of training samples cannot be achieved with just a few point clouds. On top of that, geographic generalization has yet to be thoroughly assessed, as most analyses are limited to small regions.

In this work, we aim to investigate the generalization of these models on an ultra-large scale. More concretely, we analyze ultra-large-scale 3D point clouds acquired using

The associate editor coordinating the review of this manuscript and approving it for publication was Ikramullah Lali.

airborne laser scanning (ALS) along the four provinces of Galicia in northwest Spain. Our case study covers nearly $29,557.18\ \ \mathrm{km}^2$ through $36,369 \cdot 10^6$ points. These are two orders of magnitude more in terms of surface area and one order of magnitude greater number of points than any previous work applying artificial intelligence to 3D point clouds. Furthermore, we seek quantitative insights to understand the model's results by assessing the uncertainty of the predictions.

In summary, our primary research goals are:

1) Quantify the generalization error of deep learning models on different ultra-large-scale classifications.
2) Compare deep learning and machine learning models on ultra-large-scale point-wise classification.
3) Quantify the reliability of ultra-large-scale 3D point-wise predictions through uncertainty measurement.

II. RELATED WORK

This section summarizes applications of interest and state-of-the-art deep learning methods for 3D point clouds.

A. APPLICATIONS OF 3D POINT CLOUDS

In forestry applications, 3D point clouds are used to estimate the tree stem curve and the wood volume in both terrestrial laser scanning (TLS) [3] and ALS point clouds [4]. TLS and ALS point clouds have been combined to estimate tree canopy volume with a root mean square error (RMSE) around $0.18\ \mathrm{m}^3$ [5]. Other works aim to characterize the impact of wind-induced tree movement in 3D point clouds by assessing its influence on co-registration errors [6].

Concerning agriculture, a typical task is the point-wise classification of the ground in farming fields scanned through unmanned laser scanning (ULS) [7]. Besides, crop fruit segmentation provides valuable information like crop phenotype or fruit grading [8]. When time series are available, it is possible to apply change detection methods for slope analysis of vineyards over time [9] and plant growth tracking to find efficient planting strategies [10].

In urban contexts, 3D point clouds can be used to extract relevant attributes for a geographic information system (GIS), like the slope and width of streets [11]. Roads scanned with mobile laser scanning (MLS) data are a fundamental asset that enables the accurate spatial analysis of vial infrastructure after segmenting the points conforming their surface [12]. The analysis of crossing zones regarding slope, road bump characterization, or the height difference between road and sidewalk is essential for safety analysis [13]. Further applications of urban point clouds include the segmentation of hydrants [14] and the classification of urban vegetation [15], [16].

Buildings are a frequent object in urban environments. Their classification is the first step for many engineering and architectural applications. For example, it enables roof reconstruction methods to generate an accurate 3D model of the buildings with reliable topological information [17]. It is also the first step for the generation of accurate 3D geoinformation and building information models (GeoBIM) that georeference buildings with $0.156\ \ \mathrm{m}$ positioning error [18]. In general, applications related to buildings cover a wide variety of scopes, from ancient building segmentation [19] or heritage building segmentation [20] to the advanced design of military buildings [21].

B. DEEP LEARNING AND LARGE-SCALE POINT CLOUD PROCESSING

Deep learning is a ubiquitous technology nowadays. It is used in many areas such as organ segmentation from medical image [22], fault-tolerant online multi-object tracking [23], depression detection from social networks [24], language impairment diagnosis [25], and semantic segmentation of 3D point clouds [26]. In this work, we focus on deep learning models trained on ultra-large-scale massive ALS datasets with a significant amount of missing labels. This type of dataset represents better real-world problems that often contain unlabeled points, cover a much wider area (from thousands to hundreds of thousands of square kilometers), and need to work on a wide variety of geometric representations for the same types of objects (i.e., buildings with different architecture, diverse vegetation species, etc.).

Our research can help with some important technical challenges currently open. For example, those related to forestry applications, which demand a shift from regional to real-world datasets in terms of spatial scale, with a special need for scientific studies addressing spatially extensive ALS datasets [27]. We summarize some results involving buildings and vegetation (the main objects in the ultra-large-scale ALS dataset used in this work) in Table 1. The metrics in this table will be explained later on in Section III-B1. Note that a straightforward comparison with our results is not feasible due to the novelty of the data and the tasks we are solving, given the number of points, geographic scale, and geometric diversity. However, these numbers can help contextualize our results with respect to the current state of the art.

Table 2 collects various scientific works about deep learning explicitly mentioning large-scale point clouds. In terms of the number of points, large-scale deep learning works range from 1.17 million to 9, 261 million points, while our work handles 36, 369.58 million points, i.e., one order of magnitude more than Gaydon et al., which is described as "ultra-large-scale" [45]. Regarding the area, large-scale works range from $0.31\ \mathrm{km}^2$ to $628.39\ \mathrm{km}^2$. The dataset covering a wider geographic area is the FRACTAL dataset, with a $250\ \ \mathrm{km}^2$ area. Our work uses a dataset covering Galicia, with an estimated extension of $29,557.18\ \mathrm{km}^2$ according to OpenStreetMap data [46].

Thus, our research makes a novel contribution to the state-of-the-art by applying deep learning to an ultra-large-scale dataset, one order of magnitude greater than the most significant previous work in terms of the number of points and two orders of magnitude greater in terms of area. On top of that, our work explores ultra-large-scale 3D semantic

TABLE 1. Results on different open-access datasets corresponding to different laser scanning methods. The score will be either the F1-score or intersection over union (IoU), depending on the benchmark criteria. For the Semantic3D dataset, the vegetation is the mean score for the high and low vegetation classes. For the Paris-Lille dataset, the vegetation and building classes correspond to the natural and building classes of the benchmark. For the Hessigheim 3D dataset, vegetation is the average considering low vegetation, shrub, and tree classes, while building is the mean of roof and facade classes. For the Vaihingen Semantic Labeling 3D benchmark, the vegetation score is the average of low vegetation, shrubs, and trees, while the building score is the average of the roof and facade.

Dataset	Laser scanning	Model	Vegetation (%)		Building (%)	
			F1	IoU	F1	IoU
Semantic3D (full)	TLS	ConvPoint	-	74.8	-	96.3
		PointNet++	-	58.0	-	75.9
Paris-Lille-3D	MLS	KPConv	-	91.4	-	94.0
		SFL-Net	-	91.3	-	96.4
Hessigheim 3D	ULS	KPConv	87.1	-	91.1	-
		PointNet++	59.4	-	60.9	-
Vaihingen Semantic Labeling	ALS	Hierarchical CRFs	69.1	-	75.3	-
		Multi-scale CNN	75.8	-	68.1	-
		alsNet (PointNet++)	63.2	-	70.2	-

segmentation under the unusual challenging conditions of having only 61.52% of the training points reliably labeled. This distribution represents approximately 26.62 to 38.48% less unlabeled training points than the usual scientific datasets in the context of geographic 3D point clouds.

III. METHOD

We use a workstation to prepare a massive ALS dataset with RGB color and near-infrared (NIR) data for the classification experiments and the Galician Supercomputing Center's (CESGA) FinisTerrae-III supercomputer for training and predictions. We trained two models per node, one with each GPU, using one of the available Intel Xeon Ice Lake 8352Y CPUs and allocating 123 GB of RAM at most (half of the node's real RAM). We combine SQL queries with geographic boundaries from OpenStreetMap [46] to quantify the performance of the models in different Galician provinces. Figure 1 summarizes our data, workflow, and experiments. The rest of this section describes our method in detail.

A. DATA

The PNOA-II dataset was acquired between 2015 and 2021 in Spain using different scanners. It has 20 cm error in the vertical axis and 30 cm error in the plane. We consider two different data acquisition campaigns: 1) The Galicia West campaign using a LEICA ALS60 scanner to acquire around 13,832.06 million points over an area of 12,433.30 km^2 between July 2015 and September 2015. 2) The Galicia East campaign using a LEICA ALS80 scanner to acquire 21,561.63 million points over 17,123.85 km^2 between August 2016 and February 2017. Both campaigns have a minimal point density of 0.5 pts/m^2. The average overlap between scanning strips is around 38.18%. As overlapping points are not labeled, only 61.51% provide training references.

1) GEOGRAPHIC CHARACTERISTICS

Galicia is northwest of the Iberian Peninsula, between 41°N and 44°N latitude and 9.5°W and 6.5°W longitude, and is divided into four provinces: A Coruña, Pontevedra, Lugo, and Ourense. It presents significant geographical diversity, characterized by elevations ranging from sea level to 2,123 meters in the eastern mountains and a hilly topography. Specifically, up to nine different habitat types can be clearly identified [47]: coastal and halophytic habitats, coastal dunes, freshwater, temperate heath and shrubland, sclerophyllous scrub, natural and semi-natural meadow formations, raised wetlands and mires and fens, rocky habitats and caves, and various types of forests. Galicia also includes numerous Atlantic Ocean islands, which contribute to the geographic diversity. Towns and cities tend to have low-rise buildings that rarely exceed eight floors. In rural areas, the population is dispersed with single-family homes rather than concentrated in villages. This pattern presents challenges for accurately identifying and mapping buildings, as they often blend into the surrounding environment. The land cover map in Figure 2 represents data from [48] showing the diverse landscape, including forests, agricultural zones, and urban areas scattered throughout the region.

A Coruña is bordered to the north by the Cantabrian Sea and to the west by the Atlantic Ocean. It represents 31.73% of the urban parcels and 26.10% of the rural area in Galicia. Its terrain includes a rugged and granitic coastline along the Atlantic Ocean. The altitude varies from sea level to 800 meters. It is marked by frequent, abrupt changes in elevation and gradient associated with troughs eroded along prominent joints, residual stumps of uneroded rock, and supratidal terraces, often covered in large, rounded boulders [49].

Lugo, with a short coastline, is bordered to the north by the Cantabrian Sea. Lugo represents 16.12% of the urban parcels and 20.82% of the rural area in Galicia and is marked by rolling hills and river valleys. Lugo's elevation ranges from sea level up to 1,925 meters. The eastern part of the province is rather mountainous, with the Cantabrian Mountains forming a natural boundary. Conversely, the central part is relatively flat, creating diverse landscapes across the province.

Pontevedra borders the Atlantic Ocean to the west. It represents 30.71% of the urban parcels and 14.45% of the rural area in Galicia. Its terrain varies from sea level to 1,178 meters in the inland areas. The province features a rugged coastline with numerous estuaries. Inland, the terrain is characterized by rolling hills, river valleys, and low mountains. Riverine areas support alder and willow forests, while the coastal and lower regions are dominated by shrubland and agricultural fields, including vineyards, orchards, and crops like corn and vegetables.

TABLE 2. Scientific works using deep learning and including the term "large-scale". The second number in parentheses in the area column refers to the geographic area, and it is given only when the area of the dataset is different from the geographic area (due to significant spatial overlapping over time).

Dataset	Scanning	Points (10^6)	Area (km^2)	Labeled (%)	Large-scale works
Vaihingen Semantic Labeling 3D	ALS	1.17	0.31	100.00	[28]
Oakland 3D	MLS	1.60	0.14	98.85	[29], [30]
Toronto-3D	MLS	78.30	0.37	96.70	[31], [32], [33]
Annotated DublinCity	ALS	260.00	2.00	88.14	[34]
DALES	ALS	505.00	10.00	98.48	[35], [36]
Saint-Jean	ALS	1,619.00	102.00	-	[36]
Montreal	ALS	2,439.00	123.00	-	[36]
SensatUrban	UAP	3,752.00	7.64	100.00	[37], [38], [33]
Semantic3D	TLS	4,197.00	1.67	92.59	[39], [40], [41], [37], [42]
Semantic-KITTI	MLS	5,298.00	628.39 (17.20)	98.02	[43], [44], [42], [38]
FRACTAL	ALS	9,261.00	250.00	99.40	[45]
PNOA-II Galicia	**ALS**	**36,369.58**	**29,557.18**	**61.52**	**Ours**

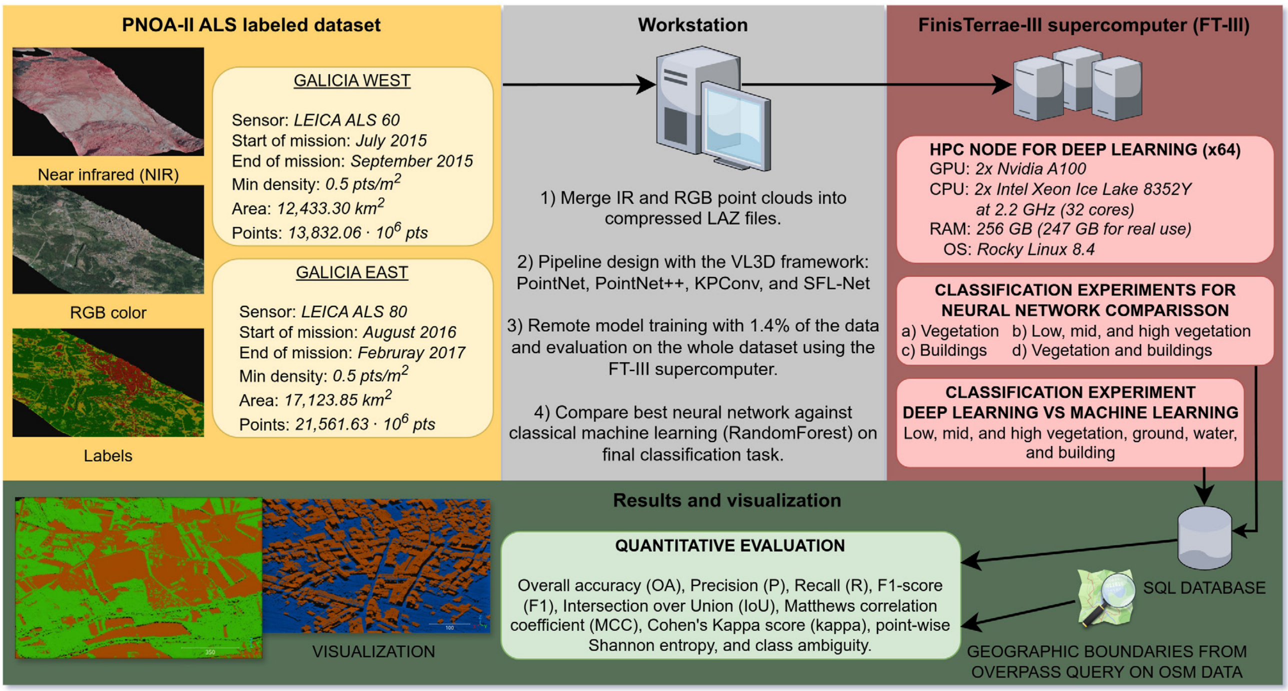


FIGURE 1. An overview of the general workflow for our experiments. First, the data is downloaded. Then, it is prepared in a local workstation and uploaded to the FinisTerrae-III supercomputer, which is used to train and evaluate the models. The results from the massive analysis are inserted into an SQL database to be properly handled. Polygonal boundaries from OpenStreetMap are used to provide region-wise evaluations. We conducted both quantitative and qualitative evaluations for all the experiments.

Ourense is distinct from the other Galician provinces due to its inland position and more continental climate, with no direct access to the coast. It represents 21.44% of the urban parcels and 25.22% of the rural area in Galicia. The terrain is characterized by rugged mountains, deep valleys, and river gorges, with elevations ranging from 39 to 2, 123 meters. The Miño River and its tributaries are significant geographical features that cut through the landscape, forming outstanding canyons.

B. EXPERIMENTS

We select one massive point cloud (with a number of points in the order of magnitude of 10^8) per province to build our training datasets. We use SQL queries to select the point clouds of a given region, maximizing the number of points for each class of interest. For multiclass classification, we prioritize point clouds with acceptable class balance, i.e., we avoid maximizing the number of samples for one class at the expense of having too few for the others. We conduct

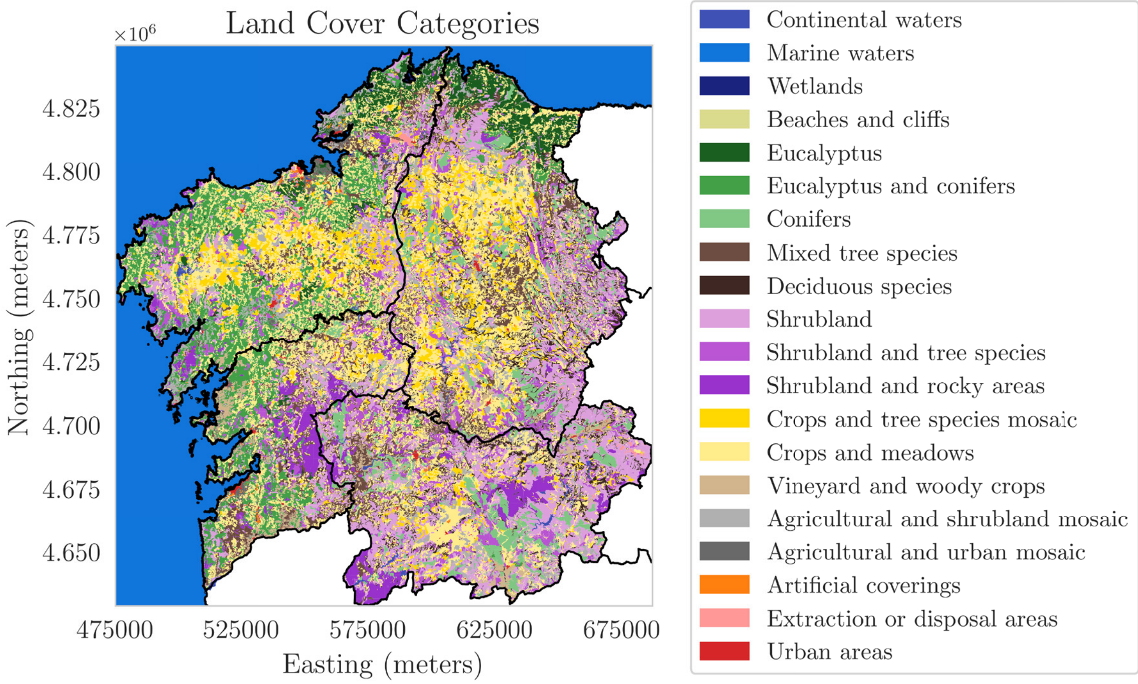


FIGURE 2. Land cover map of Galicia showing natural, agricultural, and urban areas.

preliminary experiments on four point clouds each to select the best model for the final experiments described below:

1) **Vegetation classification**. Classify the points depending on whether they are vegetation. There is no distinction between different types of vegetation. This task represents a well-balanced binary classification with training point clouds having a percentage of vegetation points ranging from 34% to 48%.
2) **Building classification**. Classify the points to determine whether they represent a building or not. There is no distinction between types of buildings. This task represents an unbalanced binary classification because the percentage of building points ranges from 1% to 2%.
3) **Low, mid, and high-vegetation classification**. Classify the points as low, mid, high-vegetation, or any other object. This task represents a multiclass classification where the classes belong to the same category.
4) **Building and vegetation classification**. Classify the points as vegetation, buildings, or any other object. This task represents a multiclass classification involving different types of objects.
5) **Full classification**. Classify the points as low, mid, or high-vegetation, ground, water, or building. We compare deep learning with machine learning through this multiclass classification.

1) VALIDATION AND ASSESSMENT

We report the Overall Accuracy (OA), Intersection over Union (IoU), Precision (P), Recall (R), and F1-score (F1) across all the Galician territory, along with two correlation metrics: Matthew's Correlation Coefficient (MCC) and Cohen's Kappa score (κ). Additionally, we use the polygonal boundaries from OpenStreetMap to compute the evaluations for each province. We also compute these metrics for each acquisition campaign and provide visualizations to support qualitative evaluation. It is worth mentioning that each evaluation metric is given twice: 1) without accounting for class distribution, i.e., all the classes are averaged independently of their number of points and 2) accounting for class distribution, i.e., the scores are aggregated considering the number of points per class (in this case, we append an "w" before the metric's name, e.g., wF1).

Note that the raw output of a neural network for c classes is a normalized probability vector $\hat{\boldsymbol{z}} \in (0, 1]^c \subset \mathbb{R}^c$ from a softmax activation function [50] Thus, we can use the entropy $\mathcal{E}$ as defined in Equation (1) based on Shannon's entropy to quantify the uncertainty as in active learning contexts [51], [52]. Alternatively, we propose the class ambiguity $\Lambda = 1 - z^* + z_*$ where z^* refers to the maximum component of the softmax vector $\hat{\boldsymbol{z}}$ and z_* refers to its second maximum component. When the difference between the two most likely predictions is small, then $\Lambda \to 1$, i.e., high uncertainty. When the difference is large, then $\Lambda \to 0$, i.e., low uncertainty.

$$\mathcal{E} = -\sum_{k=1}^{c} \hat{z}_k \log_2(\hat{z}_k) \tag{1}$$

We compute the Pearson's correlation coefficient $r \in [-1, 1] \subset \mathbb{R}$ to measure the linear relationship between the uncertainty and the F1 considering entropy and class ambiguity cuts, i.e., points with equal or lower values than the cut threshold. The ideal correlation is given by $r = -1$, implying an inverse linear relationship between the uncertainty and the F1. We also use Spearman's correlation coefficient $\rho \in [-1, 1] \subset \mathbb{R}$ to measure the monotonic relationship. The ideal correlation is given by $\rho = -1$.

C. MODELS

We use models based on the PointNet [53], PointNet++ [54], KPConv [55], and SFL-Net [56] architectures available in the open-access VirtuaLearn3D (VL3D) framework. This software allows the configuration of hyperparameters and custom architectures through JSON files.

1) NEURAL NETWORK ARCHITECTURES

First, we consider the **PointNet** model [53] as a baseline. It is based on the PointNet operator, i.e., a universal approximator for sets of points arising as the composition of three functions. First, it receives a set of points as input for a continuous component-wise function that maps each input point to a high-dimensional space. Then it computes the component-wise maximum over the input points, e.g., global max pooling. The last function is an MLP that transforms the maximum values into the final output.

We also consider the hierarchical feature extraction architecture **PointNet++** that applies the PointNet operator on consecutive receptive fields sorted from denser to less populated ones [54]. Let $\boldsymbol{P}_t = [\boldsymbol{X}_t|\boldsymbol{F}_t] \in \mathbb{R}^{m_t \times (3+n_t)}$ be the matrix representation of a neighborhood at depth t where $\boldsymbol{X}_t \in \mathbb{R}^{m_t \times 3}$ is the 3D structure space matrix (i.e., matrix of coordinates) and $\boldsymbol{F}_t \in \mathbb{R}^{m_t \times n_t}$ is the feature space matrix. The hierarchical feature extraction process begins with encoding layers and concludes with decoding layers.

The encoding layers transform $\boldsymbol{P}_t = [\boldsymbol{X}_t|\boldsymbol{F}_t]$ into $\boldsymbol{P}_{t+1} = [\boldsymbol{X}_{t+1}|\boldsymbol{F}_{t+1}] \in \mathbb{R}^{m_{t+1} \times (3+n_{t+1})}$ with $m_t > m_{t+1}$ and $n_t < n_{t+1}$, i.e., less points but more features. The decoding layers transform $\boldsymbol{P}_t = [\boldsymbol{X}_t|\boldsymbol{F}_t]$ into $\boldsymbol{P}_{t+1} = [\boldsymbol{X}_{t+1}|\boldsymbol{F}_{t+1}]$ usually to the contrary than encoding layers, i.e., with $m_t < m_{t+1}$ and $n_t > n_{t+1}$. Sampling and grouping stages connect consecutive receptive fields through the furthest point sampling (FPS) algorithm and point-wise local neighborhoods.

We obtain the best results with PointNet++ using a function with exponential weights in the encoding layers. Besides, we find that the nearest upsampling strategy works better for the decoding layers. This strategy increases the dimensionality by taking the values from the closest neighbor in the space with reduced dimensionality.

We also include the widely used **KPConv** model in our experiments [55]. It uses a kernel of $K \in \mathbb{Z}_{>0}$ points at each encoding layer. This kernel is represented by a structure space matrix $\boldsymbol{Q} \in \mathbb{R}^{K \times 3}$ and a tensor of weights $\mathcal{W} \in \mathbb{R}^{K \times D_{\text{in}} \times D_{\text{out}}}$ such that its K slices are the weights for each kernel point. The structure space of the kernel is initialized by solving the energy minimization problem in Equation (2). The kernel is assumed to be centered at zero, so $\|\boldsymbol{q}_{k*}\|^2$ represents the attractive force of the center for each k-th kernel point, and $\|\boldsymbol{q}_{l*} - \boldsymbol{q}_{k*}\|$ for $k \neq l$ represents the repulsive force that the k-th point exerts on the l-th point. The solution yields kernel points that provide good spatial coverage of a sphere's surface. The original paper solves the minimization problem using gradient descent [2]. Nonetheless, algorithms that improve the solution in the direction of the local downhill gradient at each iteration are naive approaches with poor convergence. Hence, we propose using the conjugate gradient method, which offers better convergence at the expense of requiring more memory to utilize additional information, such as the search direction and residual from previous iterations [57].

$$\boldsymbol{Q}^* = \underset{\boldsymbol{Q} \in \mathbb{R}^{K \times 3}}{\arg\min} \sum_{k=1}^{K} \left(\|\boldsymbol{q}_{k*}\|^2 + \sum_{l=1, l \neq k}^{K} \|\boldsymbol{q}_{l*} - \boldsymbol{q}_{k*}\|^{-1} \right) \tag{2}$$

After initialization, the kernel is ready to compute the convolution formalized in Equation (3). In this equation, $\sigma \in \mathbb{R}_{>0}$ governs the region of influence of the kernel; $\boldsymbol{x}_{i*} \in \mathbb{R}^3$ is the input point being convoluted; $\mathcal{N}(\boldsymbol{x}_{i*})$ is a neighborhood centered on the input point; $\boldsymbol{x}_{j*} \in \mathbb{R}^3$ represents a neighbor of the input point, $\boldsymbol{q}_{k*} \in \mathbb{R}^3$ is the k-th point of the kernel (i.e., the k-th row of the matrix $\boldsymbol{Q}$); $\boldsymbol{W}_k \in \mathbb{R}^{D_{\text{in}} \times D_{\text{out}}}$ is the k-th matrix of weights associated to the point $\boldsymbol{q}_{k*}$ (i.e., the k-th slice of the previously described tensor of weights $\mathcal{W}$); and $\boldsymbol{f}_{j*} \in \mathbb{R}^{D_{\text{in}}}$ is the features vector from the corresponding neighbor of the input point. The result of this convolution is a new D_{out}-dimensional feature vector. Contrary to the original paper, we found that using a mean-based upsampling worked better than nearest-neighbor upsampling.

$$\begin{aligned} &\left(\boldsymbol{P} * \boldsymbol{Q}\right)(\boldsymbol{x}_{i*}) \\ &= \sum_{\boldsymbol{x}_{j*} \in \mathcal{N}(\boldsymbol{x}_{i*})} \sum_{k=1}^{K} \max\left\{0, 1 - \frac{\|\boldsymbol{x}_{j*} - \boldsymbol{x}_{i*} - \boldsymbol{q}_{k*}\|}{\sigma}\right\} \boldsymbol{W}_k^{\intercal} \boldsymbol{f}_{j*} \end{aligned} \tag{3}$$

Finally, we use a model based on **SFL-Net**, which involves modifying two primary components of KPConv. First, the convolution operator is simplified to work with a vector of weights per kernel point. This simplification is achieved using Equation (4). In this equation, there is a single matrix of weights $\boldsymbol{W} \in \mathbb{R}^{D_{\text{in}} \times D_{\text{out}}}$ and a diagonal matrix $\boldsymbol{A}(\boldsymbol{x}_{i*}, \boldsymbol{x}_{j*}) \in \mathbb{R}^{D_{\text{in}} \times D_{\text{in}}}$ that is the summation of the scaled contributions for each neighbor following Equation (5), where $\boldsymbol{a}_k \in \mathbb{R}^{D_{\text{in}}}$ represents the scale factors for the k-th kernel point. The residual block uses an hourglass operator that transforms the features in two steps as described by [56]. It uses one unbiased MLP to reduce the dimensionality of the feature space from $D_{\text{in}} \in \mathbb{Z}_{>0}$ to $D_h \in \mathbb{Z}_{>0}$ and then another to transform it to the output feature space with $D_{\text{out}} \in \mathbb{Z}_{>0}$ features, where typically

$D_h < D_{\text{out}}$. The output of an hourglass block for m points is given by $\sigma_2(\sigma_1(\boldsymbol{F\Omega}_1)\boldsymbol{\Omega}_2) \in \mathbb{R}^{m\times D_{\text{out}}}$, where $\boldsymbol{F} \in \mathbb{R}^{m\times D_{\text{in}}}$ is the input feature space, $\boldsymbol{\Omega}_1 \in \mathbb{R}^{D_{\text{in}}\times D_h}$ are the weights for the first unbiased MLP, σ_1 represents its activation function (typically a ReLU), $\boldsymbol{\Omega}_2 \in \mathbb{R}^{D_h\times D_{\text{out}}}$ are the weights for the second unbiased MLP, and σ_2 its activation function.

$$\Big(\boldsymbol{P} * \boldsymbol{Q}\Big)(\boldsymbol{x}_{i*}) = \sum_{x_{j*}\in\mathcal{N}(\boldsymbol{x}_{i*})} \Big(\boldsymbol{A}(\boldsymbol{x}_{i*}, \boldsymbol{x}_{j*})\boldsymbol{W}\Big)^{\intercal} \boldsymbol{f}_{j*} \tag{4}$$

$$\begin{aligned}&\boldsymbol{A}(\boldsymbol{x}_{i*}, \boldsymbol{x}_{j*})\\ &= \operatorname{diag}\left(\sum_{k=1}^{K} \max\left\{0,\, 1 - \frac{\|\boldsymbol{x}_{j*} - \boldsymbol{x}_{i*} - \boldsymbol{q}_{k*}\|}{\sigma}\right\}\boldsymbol{a}_{k*}\right)\end{aligned} \tag{5}$$

2) RANDOM FOREST

For the Random Forest model, we consider linearity, planarity, surface variation, verticality, and anisotropy features in spherical neighborhoods with radii of 10, 30, and 60 meters. We also include the floor distance (vertical distance with the lowest point in the neighborhood) and ceil distance (vertical distance with the highest point in the neighborhood) computed in a rectangular neighborhood with an edge size of 200 meters and the mean of the NIR and RGB components in a spherical neighborhood with 30 meters radius. These features are used in other works like [58], [59], and [60]. We also train an alternative Random Forest that incorporates floor and ceiling distances in rectangular neighborhoods with an edge size of 100 meters and the mean of the NIR and RGB components in a second spherical neighborhood with a radius of 90 meters.

We conduct hyperparameter tuning on the Random Forest models to find an appropriate configuration, exploring the maximum depth, minimum number of samples required to split a node, and minimum number of samples required to accept a node as a leaf. The best Random Forest in our experiments is the one using all the features (including the extra height and mean NIR and RGB features) with a maximum depth of 30 for each decision tree, a minimum of 32 samples to split a node, and a minimum of 8 samples to accept a leaf.

D. SEMI-SUPERVISED TRAINING PROCEDURE

Although, on average, more than 38% of the points do not belong to a reliable class (see Section III-A), they must be fed into the neural network because they yield relevant geometric information. Removing them is unacceptable because we cannot assume which points are noisy or come from overlapping regions beforehand. Thus, all models must be formalized as multiclass neural networks supporting an extra class for overlapping and noisy points.

For this reason, we use a softmax function in the final layer and the categorical cross-entropy loss described in Equation (6), where $\boldsymbol{y} \in \{0, 1\}^c$ and $\hat{\boldsymbol{z}} \in (0, 1]^c \subset \mathbb{R}^c$ are the vector of reference and predicted classes, respectively. Any component of the reference vector $\boldsymbol{y}$ is either 0 or 1 as it is the one-hot encoded representation of the reference classes [50]. Any component of the predictions vector is inside the real $(0, 1)$ interval due to the softmax function of the last layer. Moreover, $\boldsymbol{\omega}$ is the real vector of parameters governing the model $\hat{\boldsymbol{z}}(\boldsymbol{\omega})$, leading to the training gradient $\nabla_{\boldsymbol{\omega}}\mathcal{L}$ used to update the model's parameters such that $\boldsymbol{\omega}_{t+1} = \boldsymbol{\omega}_t - \alpha\nabla_{\boldsymbol{\omega}}\mathcal{L}$ (with $\alpha \in \mathbb{R}_{>0}$ being the learning rate). Unfortunately, this standard procedure updates the model's weights even when the reference vector $\boldsymbol{y}$ corresponds to the class that must not directly influence the gradient.

$$\mathcal{L}(\boldsymbol{y}, \hat{\boldsymbol{z}}) = -\sum_{k=1}^{c} y_k \log(\hat{z}_k)\ , \quad \nabla_{\boldsymbol{\omega}}\mathcal{L} = -\sum_{k=1}^{c} \frac{y_k \nabla_{\boldsymbol{\omega}}\hat{z}_k}{\hat{z}_k} \tag{6}$$

To solve this issue, we redefine the loss function to be $\mathcal{L}'(\boldsymbol{y}, \hat{\boldsymbol{z}}) = \langle\boldsymbol{\theta}, \boldsymbol{y}\rangle\mathcal{L}(\boldsymbol{y}, \hat{\boldsymbol{z}})$ where $\langle\cdot,\cdot\rangle$ represents the dot product and $\boldsymbol{\theta} \in \mathbb{R}^c_{\geq 0}$ is the vector of class weights. Consequently, the gradient will be formalized by Equation (7). If we define $\boldsymbol{\theta}$ subject to $\theta_k = 0$ when k is the index of the special class, we have a class-weighted categorical cross-entropy loss function that prevents updates from not reliably labeled points.

$$\nabla_{\boldsymbol{\omega}}\mathcal{L}' = -\big\langle\boldsymbol{\theta}, \boldsymbol{y}\big\rangle \sum_{k=1}^{c} \frac{y_k \nabla_{\boldsymbol{\omega}}\hat{z}_k}{\hat{z}_k} \tag{7}$$

In the first four training processes, we train each model using a point cloud from a different province. In the fifth training process, we revisit the first point cloud. We lower the initial learning rate on consecutive training processes to avoid catastrophic forgetting, as recommended in continual learning contexts [61]. The precise configuration of our training procedures for each model is available in the project's public repository (see Section VII).

Concerning semi-supervised Random Forest training, our strategy is straightforward. First, we include all the points when computing the geometric features. Then we consider only those points with a reliable label as training samples, i.e., ignoring noise and overlapping.

IV. RESULTS

This section presents the results of the preliminary experiments for model selection, the results of the five extensive experiments, and the uncertainty analysis.

A. MODEL SELECTION RESULTS

Table 3 summarizes the preliminary results. The SFL-Net models yield the best scores in three preliminary experiments, while KPConv is the best for binary building classification. Thus, we select the SFL-Net model, trained on point clouds with NIR and RGB features, for the final extensive experiments on the entire dataset.

B. VEGETATION CLASSIFICATION RESULTS

Table 4 summarizes the results of SFL-Net for binary vegetation classification, while Figure 3 provides visual inspection.

TABLE 3. The results of the preliminary experiments on binary and multiclass classification tasks in terms of F1-score (F1) and Matthew's Correlation Coefficient (MCC) for each model. The best result for each task is highlighted in bold.

Neural Network	Input	Vegetation F1 (%)	Vegetation MCC (%)	Building F1 (%)	Building MCC (%)	LMH vegetation F1 (%)	LMH vegetation MCC (%)	Build. and Veg. F1 (%)	Build. and Veg. MCC (%)
PointNet	X	83.73	67.50	75.98	57.66	64.92	64.95	74.24	69.41
	X, NIR	83.69	67.75	68.68	40.95	63.59	66.75	70.64	67.80
	X, RGB	85.58	71.69	89.36	79.15	67.95	69.59	83.40	70.90
	X, NIR, RGB	84.85	71.09	90.10	80.65	65.13	67.78	82.81	68.56
PointNet++	X	84.03	68.39	85.07	70.99	60.95	63.39	80.23	69.81
	X, NIR	83.92	68.07	87.88	76.80	59.35	64.86	79.02	69.75
	X, RGB	84.79	69.93	90.95	82.57	61.47	65.96	82.49	70.81
	X, NIR, RGB	84.79	69.93	90.01	80.12	61.06	65.20	83.32	71.10
KPConv	X	85.58	72.12	92.98	86.23	68.75	68.10	84.24	73.69
	X, NIR	85.84	71.82	92.18	84.54	68.98	70.75	86.25	75.04
	X, RGB	86.22	72.81	**95.21**	**90.47**	69.15	68.64	87.55	75.56
	X, NIR, RGB	86.12	72.44	94.61	89.40	69.34	70.67	87.95	76.12
SFL-Net	X	86.07	72.50	92.64	85.29	69.63	70.99	86.71	75.14
	X, NIR	86.01	72.23	91.79	83.60	68.76	69.40	85.55	73.94
	X, RGB	86.55	**73.45**	94.16	88.58	69.87	71.60	88.24	76.59
	X, NIR, RGB	**86.57**	73.34	94.38	88.97	**71.00**	**71.82**	**88.36**	**76.73**

TABLE 4. Results of the SFL-Net model on binary vegetation classification using (x, y, z), near-infrared, and RGB data.

Region	Scores (%) OA	P	R	F1	IoU	wP	wR	wF1	wIoU	MCC	κ
A Coruña	88.24	87.94	88.12	87.97	78.65	88.37	88.24	88.25	79.08	76.05	75.95
Lugo	84.43	84.36	84.25	84.08	72.87	84.83	84.43	84.41	73.37	68.60	68.24
Pontevedra	**88.37**	**88.05**	**88.28**	**88.11**	**78.86**	**88.50**	**88.37**	**88.38**	**79.29**	**76.33**	**76.24**
Ourense	83.55	83.49	83.39	83.25	71.67	83.86	83.55	83.52	72.06	66.87	66.58
Galicia West	88.29	87.99	88.19	88.03	78.74	88.43	88.29	88.31	79.17	76.17	76.08
Galicia East	84.00	83.93	83.82	83.67	72.27	84.35	84.00	83.97	72.72	67.74	67.42
Galicia	**85.45**	**85.28**	**85.31**	**85.13**	**74.42**	**85.76**	**85.45**	**85.45**	**74.89**	**70.58**	**70.32**

TABLE 5. Class-wise results of the SFL-Net model on low, mid, and high-vegetation classification using (x, y, z), near-infrared, and RGB data.

Region	Prec. per-class (%) Low	Mid	High	Recall per-class (%) Low	Mid	High	F1 per-class (%) Low	Mid	High	IoU per-class (%) Low	Mid	High
A Coruña	60.83	57.12	**95.65**	48.11	32.61	97.10	53.30	41.27	96.34	36.78	26.16	92.99
Lugo	51.37	**60.68**	94.17	36.13	31.74	97.79	41.79	41.40	95.90	26.69	26.43	92.20
Pontevedra	**61.08**	57.70	95.55	**49.88**	**34.34**	97.75	**54.77**	**42.94**	**96.63**	**37.98**	**27.45**	**93.50**
Ourense	51.86	59.22	93.61	36.48	31.46	**98.11**	42.49	40.95	95.80	27.32	25.86	91.98
Galicia West	60.94	57.37	95.61	48.89	33.37	97.39	53.95	42.00	96.47	37.31	26.73	93.21
Galicia East	51.61	59.95	93.89	36.31	31.60	97.95	42.14	41.17	95.85	27.00	26.15	92.09
Galicia	**54.55**	**59.27**	**94.57**	**40.92**	**32.27**	**97.86**	**46.25**	**41.55**	**96.17**	**30.60**	**26.45**	**92.67**

TABLE 6. Results of the SFL-Net model on binary building classification using (x, y, z), near-infrared, and RGB data.

Region	Scores (%) OA	P	R	F1	IoU	wP	wR	wF1	wIoU	MCC	κ
A Coruña	99.45	86.71	93.31	88.34	82.79	99.69	99.45	99.56	99.18	78.10	76.76
Lugo	99.60	83.40	89.45	84.60	77.76	**99.80**	99.60	99.69	99.43	70.81	69.26
Pontevedra	99.65	**91.51**	**93.61**	**92.21**	**86.93**	99.69	99.65	99.66	99.38	**84.76**	**84.42**
Ourense	**99.71**	85.42	90.80	86.88	80.37	99.79	**99.71**	**99.74**	**99.54**	74.88	73.78
Galicia West	99.54	88.82	93.44	90.04	84.61	99.69	99.54	99.60	99.27	81.03	80.12
Galicia East	99.66	84.40	90.12	85.73	79.05	99.79	99.66	99.72	99.48	72.82	71.50
Galicia	**99.62**	**86.44**	**91.16**	**87.48**	**81.26**	**99.75**	**99.62**	**99.67**	**99.41**	**76.13**	**75.00**

Note that ground-level vegetation seems problematic, while high-vegetation (e.g., trees) is accurately classified. Further results, including those for low, mid, and high-vegetation types, as well as the uncertainty assessment presented

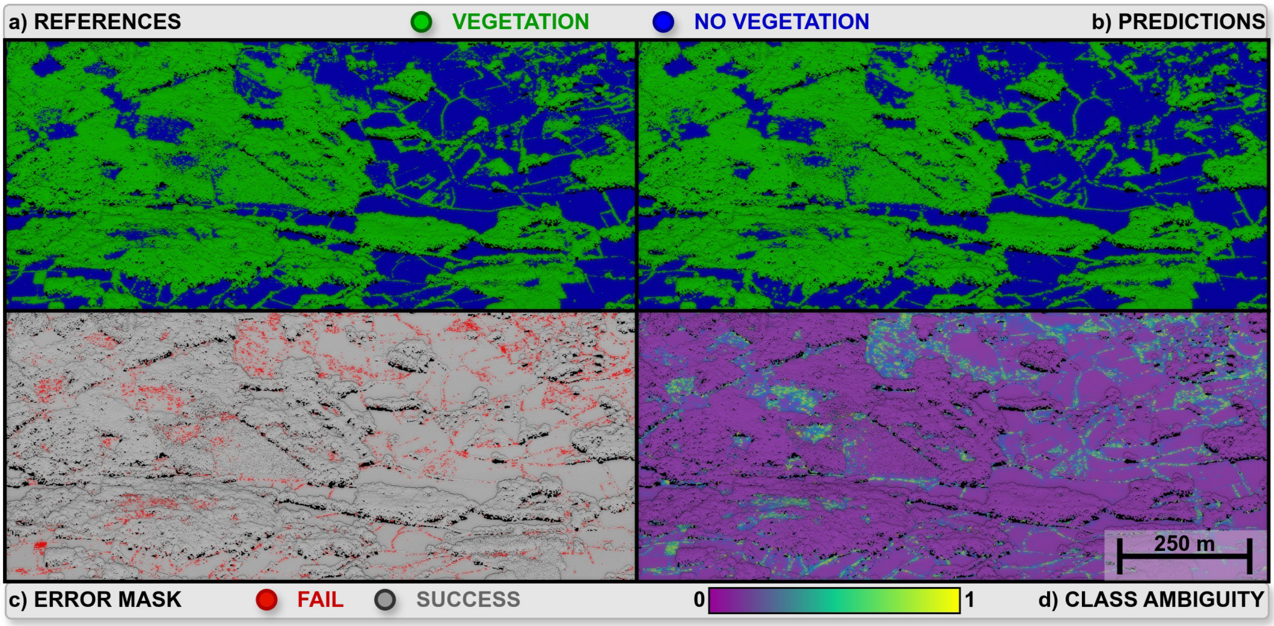


FIGURE 3. Visualization of the results yielded by the SFL-Net model trained on (x, y, z), near-infrared, and RGB data for binary vegetation classification. The labeled reference data is shown in a), the classification yielded by the model in b), c) shows an error mask that colors the discrepancies between the reference labels and the predicted ones in red, and d) represents the class ambiguity used as an uncertainty measurement. The represented point cloud is centered around 42°46′59.7″N 7°51′32.9″W, near the Galician river known as Fradal.

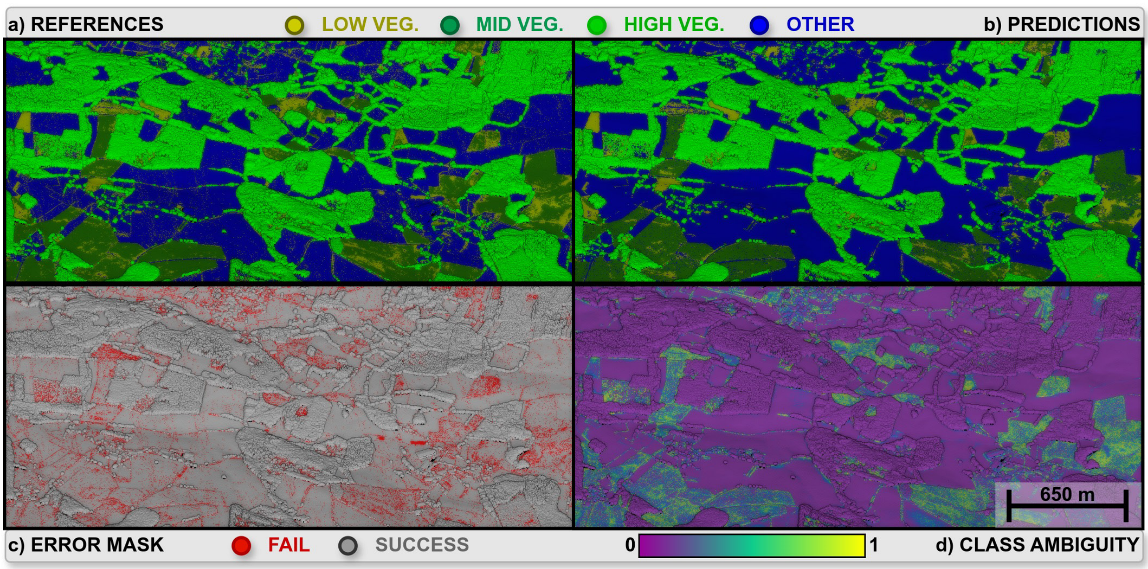


FIGURE 4. Visualization of the results yielded by the SFL-Net model trained on (x, y, z), near-infrared, and RGB data for low, mid, and high-vegetation classification. The labeled reference data is shown in a), the classification yielded by the model in b), c) shows an error mask that colors the discrepancies between the reference labels and the predicted ones in red, and d) represents the class ambiguity used as an uncertainty measurement. The represented point cloud is centered around 43°03′48.2″N 7°23′27.3″W, surrounded by the villages of A Meda, Córneas, and Vilafrío in Lugo, Galicia.

in Section IV-H, quantify the differences in performance depending on the vegetation type.

C. LOW, MID, AND HIGH-VEGETATION CLASSIFICATION RESULTS

Table 5 helps to understand the problem illustrated for the binary case in Figure 3. More concretely, SFL-Net achieves a 49.9% greater F1 score for high-vegetation compared to low-vegetation, and a 54.6% greater F1 score compared to mid-vegetation. Figure 4 provides a visual representation of the results, coloring the points by vegetation type. Note that the error is concentrated in low and mid-vegetation regions, while high-vegetation regions are accurately classified. These results show that we can expect outstanding high-vegetation

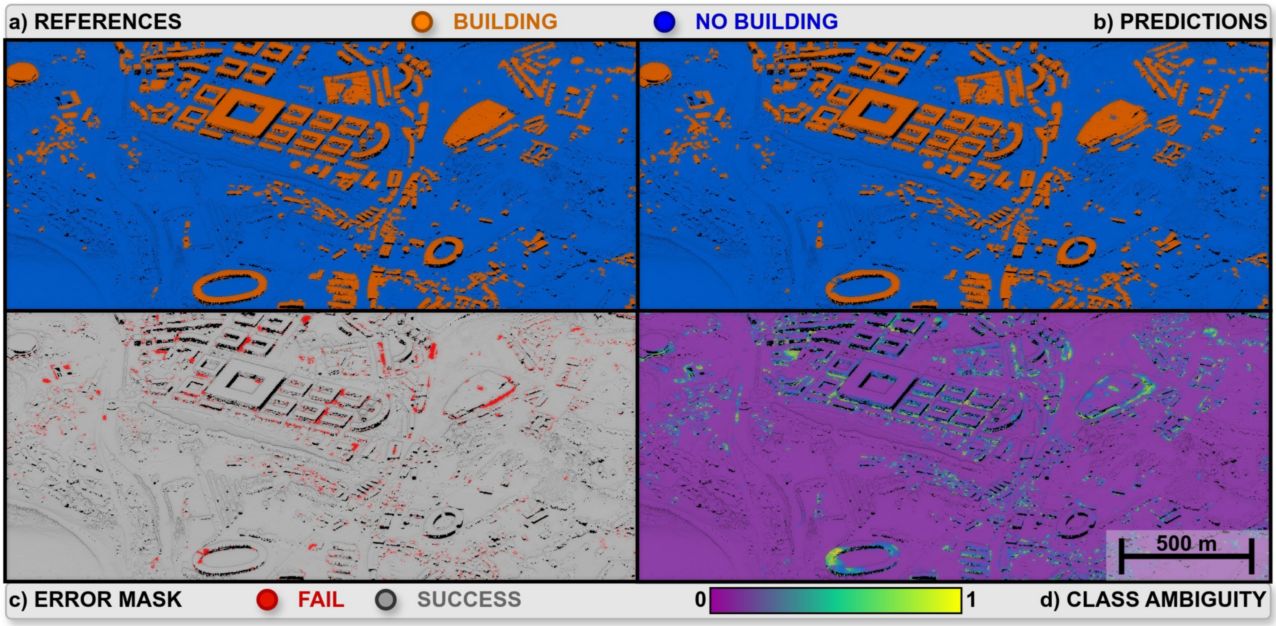


FIGURE 5. Visualization of the results yielded by the SFL-Net model trained on (x, y, z), near-infrared, and RGB data for the task of binary building classification. The labeled reference data is shown in a), the classification yielded by the model in b), c) shows an error mask that colors the discrepancies between the reference labels and the predicted ones in red, and d) represents the class ambiguity used as an uncertainty measurement. The represented point cloud is centered around 42°53′06.1″N 8°31′14.3″W, in the eastern region of the Galician city Santiago de Compostela.

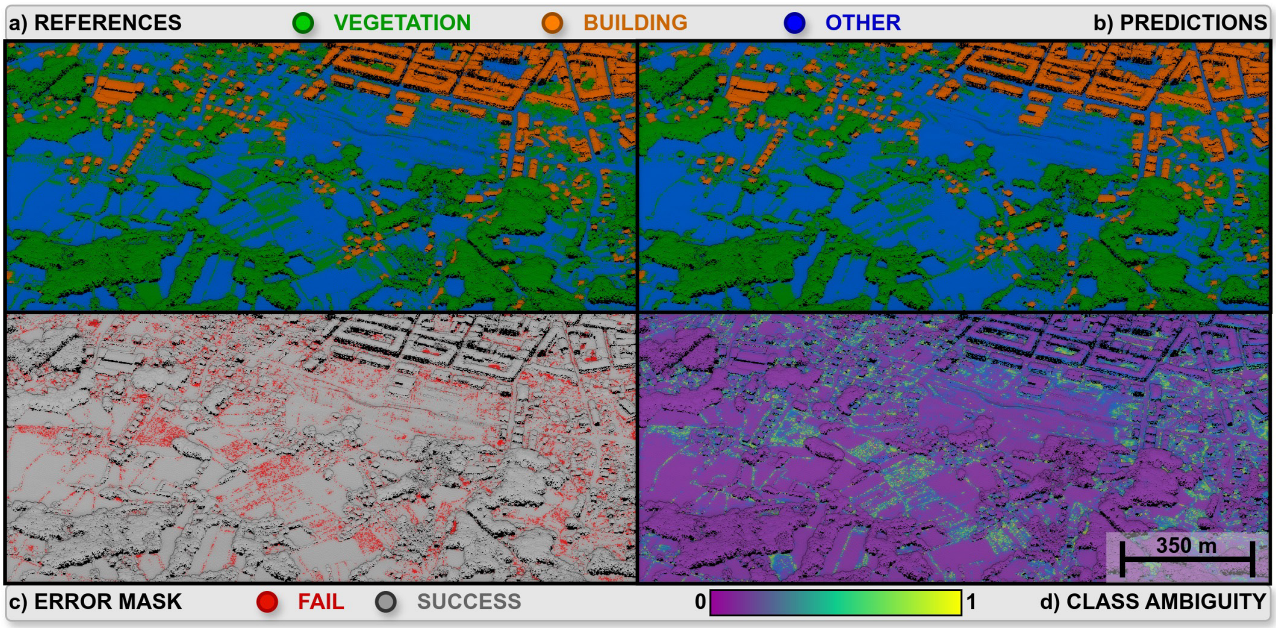


FIGURE 6. Visualization of the results yielded by the SFL-Net model trained on (x, y, z), near-infrared, and RGB data for building and vegetation classification. The labeled reference data is shown in a), the classification yielded by the model in b), c) shows an error mask that colors the discrepancies between the reference labels and the predicted ones in red, and d) represents the class ambiguity used as an uncertainty measurement. The representation of the point cloud is centered around 42°39′21.2″N 8°06′21.3″W, near the Galician village known as Lalín.

classification on ultra-large-scale datasets (29, 557 km^2) even under poor resolution conditions (only 0.5 pts/m^2 guaranteed, with approximately 1.2 pts/m^2 on average).

D. BUILDING CLASSIFICATION RESULTS

Table 6 summarizes the results on binary building classification. The OA reveals that more than 99% of the points are correctly classified. However, this classification presents a significantly unbalanced class distribution (approximately 32 times as many non-building as building points). The non-weighted F1 is 12.19% lower than the weighted F1, indicating that while most buildings are correctly classified, there are still a few errors. Some of these errors are due to incorrectly labeled points in the reference data.

Recall and precision can give us further insights because the former is 2.1% greater than the latter in the best case.This difference implies that the error in building classification is slightly biased towards false positives rather than missing building points (false negatives). Figure 5 shows that most buildings are correctly detected, but some present missing parts. Generally, the errors occur within the boundaries of the buildings. More details about this are given in Section IV-H.

TABLE 7. Computational benchmarking for the deep learning models with NIR and RGB data trained in the building and vegetation classification task. The epoch time represents the runtime of a single training epoch in seconds. The RAM peak represents the maximum used RAM during the entire training process.

Model	**Epoch time** (s)	**RAM peak** (GB)
PointNet	255	51.72
PointNet++	33	72.01
KPConv	126	50.87
SFL-Net	284	56.58

E. BUILDING AND VEGETATION CLASSIFICATION RESULTS

Table 7 quantifies the runtime for each training epoch on the GPU and the maximum peak of RAM during the training process. The PointNet architecture was configured to work with 40,000 rectangular neighborhoods with a side length of 100 m. We did this to study whether a simpler and older model that does not rely on hierarchical geometric representations for feature extraction could improve its results by increasing its receptive field. These reasons explain its high runtime. The PointNet++ model considered only 20,000 rectangular neighborhoods, and it is much faster than PointNet because it reduces the number of points in the hierarchical representations while increasing the dimensionality of the hidden feature space. To the contrary, PointNet increases the dimensionality of the feature space while keeping the number of points constant. Note that the SFL-Net model is more than twice as slow as KPConv due to the spectral norm-based regularization of the hourglass weights during training. The regularization strategy imposes a high computational cost because it requires computing a singular value decomposition for each hourglass. This performance difference affects the model only during training, since spectral regularization is not computed in the forward pass.

Table 8 quantifies the class-wise results on building and vegetation classification. The difference between the class-wise F1 for vegetation and the F1 of the binary vegetation classification is barely around 0.61%, suggesting that the multiclass model classifies the vegetation almost as the binary one. This difference is approximately 7.51% for buildings, suggesting that a binary classification model might be preferred for the semantic segmentation of buildings. The results are visually represented in the example of Figure 5. The errors for vegetation are generally concentrated in low-vegetation regions, as in other vegetation experiments. The buildings are still detected despite some missing parts and misclassifications in their boundaries.

F. FULL CLASSIFICATION RESULTS

The results of SFL-Net are compared to those of the Random Forest model, representing a classical machine learning approach. Considering the F1, SFL-Net scores 15.44% better than Random Forest. Regarding the number of correctly classified points, SFL-Net presents an OA 7.06% better. The neural network outperforms the Random Forest across all considered evaluation metrics. The results of the deep learning model are depicted in Figure 7. As in previous experiments, high-uncertainty regions are often located in low-vegetation areas, especially those with sparse vegetation. There are errors concentrated in mid-vegetation areas that overlap with low-vegetation areas. The model is also biased toward ground predictions in regions with mixed terrain and ground-level vegetation labels.

G. QUANTITATIVE CONTEXTUALIZATION

The results from our experiments can be quantitatively contextualized with respect to the state of the art. A direct and fair comparison is not possible due to the large gaps in geographic area, geometric diversity, underlying class semantics, and the number of points between standard and ultra-large-scale datasets. Notwithstanding, we can compare our ultra-large-scale ALS models with those in standard benchmarks to contextualize our results. Note that direct comparisons will be possible in the future, as new ultra-large-scale works can be compared against the results in this paper.

Compared to the MLS dataset of Table 1, which presents the highest IoU for vegetation, our binary vegetation model is 7% IoU below. Considering the ULS dataset, which is closer to ALS than MLS, it is paired closely to ours with a minor difference of 2% F1. Compared to the ALS case, our ultra-large-scale model is 9.3% higher in F1. This confirms that the results of ultra-large-scale ALS models do not outperform those of MLS models, likely because the latter offer more accurate 3D geometric representations. However, compared to standard ULS and ALS datasets, it is possible to achieve state-of-the-art results on ultra-large-scale vegetation classification. This is a useful result, especially for the forestry community, which can expect good generalization even in national-scale models with significant species diversity and low-resolution ALS data.

Compared to the MLS dataset of Table 1, which achieves the highest IoU for buildings, our binary building model is 15.1% IoU below. Considering the ULS dataset, they are

TABLE 8. Class-wise results of the SFL-Net model on building and vegetation classification using (x, y, z), near-infrared, and RGB data.

Region	Prec. per-class (%)		Recall per-class (%)		F1 per-class (%)		IoU per-class (%)	
	Building	Veget.	Building	Veget.	Building	Veget.	Building	Veget.
A Coruña	75.23	93.20	89.35	82.84	79.58	87.60	68.86	78.15
Lugo	76.17	92.70	82.53	75.08	77.03	82.82	64.84	70.99
Pontevedra	**82.68**	**93.40**	**89.36**	**84.15**	**85.48**	**88.50**	**75.22**	**79.45**
Ourense	76.51	92.09	84.65	74.47	79.09	82.12	66.99	70.14
Galicia West	78.50	93.29	89.35	83.42	82.17	88.00	71.65	78.72
Galicia East	76.34	92.40	83.58	74.77	78.05	82.47	65.90	70.57
Galicia	**78.10**	**92.80**	**85.36**	**77.92**	**79.97**	**84.52**	**68.51**	**73.57**

TABLE 9. Results of the Random Forest and SFL-Net models on the full classification task using (x, y, z), near-infrared, and RGB data.

Region	Random Forest scores (%)							SFL-Net scores (%)						
	OA	P	R	F1	IoU	MCC	κ	OA	P	R	F1	IoU	MCC	κ
A Coruña	75.67	62.86	56.73	55.77	46.59	63.27	61.51	83.17	75.73	**71.61**	71.77	61.58	74.80	74.30
Lugo	72.10	54.52	51.05	47.79	38.61	56.87	54.80	78.69	67.22	63.02	62.58	52.26	67.34	66.35
Pontevedra	75.52	62.73	55.85	55.49	46.20	63.13	61.43	**83.63**	**76.48**	71.37	**72.63**	**62.45**	**75.52**	**75.18**
Ourense	71.28	57.11	53.90	51.07	41.46	55.67	53.49	78.09	69.74	65.87	65.75	55.05	66.40	65.52
Galicia West	75.61	62.80	56.34	55.65	46.42	63.21	61.48	83.37	76.06	71.51	72.15	61.96	75.12	74.69
Galicia East	71.69	55.80	52.45	49.41	40.02	56.28	54.15	78.39	68.47	64.43	64.15	53.64	66.88	65.94
Galicia	73.12	57.75	53.00	50.86	41.61	58.90	56.88	**80.18**	**70.60**	**66.09**	**66.30**	**56.06**	**69.87**	**69.10**

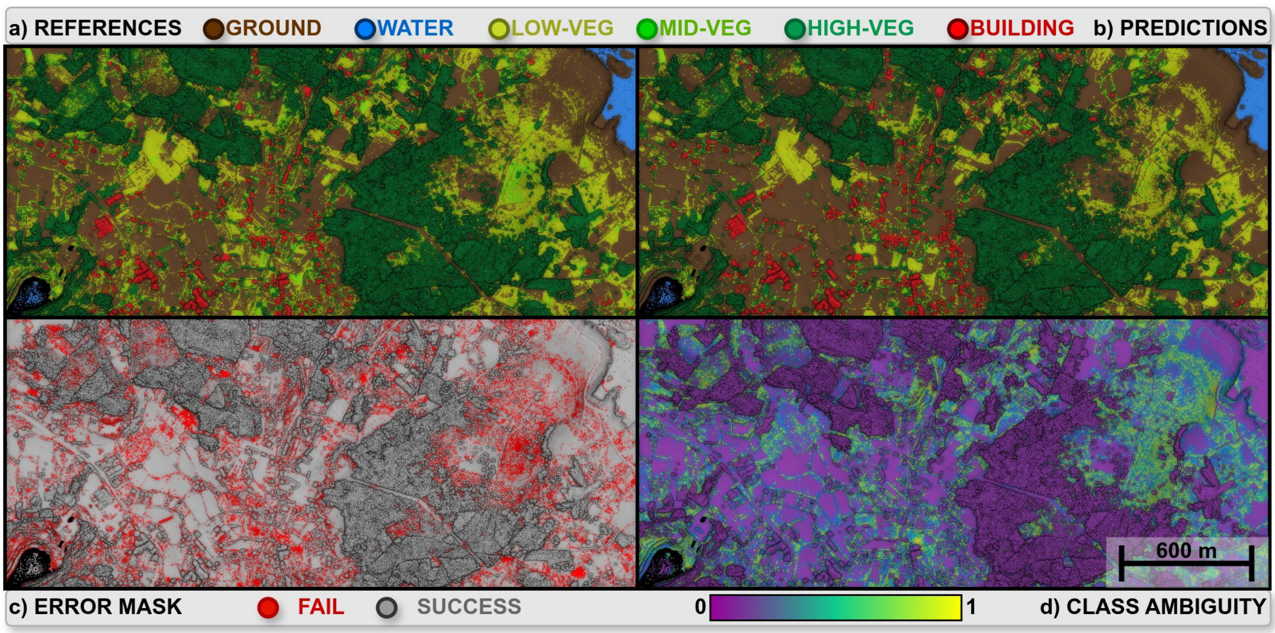


FIGURE 7. Visualization of the results yielded by the SFL-Net model trained on (x, y, z), near-infrared, and RGB data for the full classification task. The labeled reference data is shown in a), the classification yielded by the model in b), c) shows an error mask that colors the discrepancies between the reference labels and the predicted ones in red, and d) represents the class ambiguity used as an uncertainty measurement. The represented point cloud is centered around 43°18′34.6″N 8°31′24.2″W, near the Galician village known as Arteixo.

much closer, with the ULS model being 3.5% better in terms of F1. Compared to the ALS case, our ultra-large-scale model is 11.6% higher in F1. As with vegetation, it is possible to achieve state-of-the-art results compared to standard ALS study cases when working with ultra-large-scale datasets. This result validates the generalization of deep learning workflows for ultra-large-scale territorial and construction studies. Note that the good results achievable with more accurate 3D representations, such as those obtained by MLS models, cannot be matched with ALS data alone, even when training on a diverse dataset spanning different regions of the territory.

H. UNCERTAINTY ASSESSMENT AND PROBLEMS

Figure 8 summarizes the uncertainty assessment. In general, class ambiguity can be used to quantify the classification reliability when applying the model to unseen data. Quantitatively, Pearson's r_Λ and Spearman's ρ_Λ correlation coefficients are close to -1, suggesting strong inverse linear and monotonically decreasing correlations between the class ambiguity cut and the F1, respectively. The main exception is the experiment involving separated classes for low, mid, and high-vegetation. In this case, both correlation coefficients are closer to zero than -1, likely due to the problematic labeling of near-ground vegetation points, the 20 cm of vertical error, and the low resolution.

We also compute the Pearson's correlation coefficient for entropy cuts $r_\mathcal{E}$ and the Spearman's correlation coefficient $\rho_\mathcal{E}$. For binary vegetation classification, we have $r_\mathcal{E} = -0.897$ and $\rho_\mathcal{E} = -1$; for low, mid, and high-vegetation $r_\mathcal{E} = -0.015$ and $\rho_\mathcal{E} = -0.055$; for binary building classification $r_\mathcal{E} = -0.884$ and $\rho_\mathcal{E} = -1$; and for building and vegetation classification $r_\mathcal{E} = -0.967$ and $\rho_\mathcal{E} = -1$. The class ambiguity consistently outperforms entropy. On top of that, the correlation between the class ambiguity cut and the F1 is 15 times stronger in terms of Pearson's correlation and 2.76 times stronger in terms of Spearman's correlation for the low, mid, and high-vegetation cases. These results suggest that class ambiguity is more effective than entropy for point-wise uncertainty quantification.

We hold that the results of our uncertainty assessment are quantitatively reliable because the p-values associated with the measured Pearson's and Spearman's correlation coefficients are generally small, i.e., their orders of magnitude range between 10^{-4} and 10^{-64}. The only abnormally high p-values correspond to the problematic classifications of low, mid, and high vegetation due to the aleatoric uncertainty induced by near-ground vegetation points. These p-values are 0.965 (Pearson) and 0.873 (Spearman) for the entropy and 0.532 (Pearson) and 0.676 (Spearman) for the class ambiguity.

Some of the previously mentioned issues with vegetation can be understood by looking at Figure 9. We can observe the errors in low-vegetation regions that explain the significant difference of 49.92% in F1 between high-vegetation (96.17% F1) and low-vegetation (46.25% F1) areas. This explanation is consistent with the technical details because, with an error of 20 cm in the vertical axis, it is difficult to distinguish low-vegetation (e.g., grass) from the ground (e.g., soil). Moreover, the extra features (NIR or RGB) might sometimes be missing, as illustrated in Figure 10, where a region lacks RGB data. It can be seen that many buildings are incorrectly detected in this region, and the model's uncertainty increases due to the lack of color features.

V. DISCUSSION AND FUTURE WORK

This section discusses the comparison between deep and machine learning, provides insights into handling classification uncertainty on unseen data, and outlines future work.

A. DEEP LEARNING VS MACHINE LEARNING

The results in Table 9) show that deep learning outperforms machine learning for all metrics in our experiments (e.g., 15.44% better F1). In contrast to deep learning, machine learning typically requires hand-crafted feature selection. Additionally, we achieve the best results by increasing the number of features and using deeper trees, which requires more manual work hours and higher computational costs. For these reasons, we argue that deep learning should be preferred over machine learning for 3D semantic segmentation of ultra-large-scale ALS point clouds.

B. HANDLING THE UNCERTAINTY

The uncertainty assessment in our experiments (see Section IV-H) shows that it is possible to reliably quantify the uncertainty associated with each point-wise prediction on unseen data. In other words, we can distinguish reliable predictions from doubtful or problematic ones by computing the point-wise entropy or the class ambiguity. This quantification is fundamental because it enables the decision-maker to determine which regions and information can be trusted and which are more likely to be incorrect or unreliable.

After comparing the point-wise entropy typically used in active learning contexts [51], [52] with the class ambiguity, we found that the latter is more reliable due to 1) its stronger linear and monotonic correlations in the general case and 2) its 15 times better Pearson's correlation coefficient and 2.76 times better Spearman's correlation coefficient for the challenging case of low, mid, and high-vegetation. Consequently, class ambiguity can improve the effectiveness of active learning in 3D point clouds compared to point-wise entropy.

The uncertainty analysis conducted on an ultra-large-scale dataset confirms that active and weakly supervised learning strategies can be successfully applied to ultra-large-scale classification tasks in 3D point clouds. The class ambiguity introduced in this paper has many potential applications in different scenarios. In active learning loops, class ambiguity can be the uncertainty measurement guiding the selection of new samples for manual annotation. In weakly supervised learning contexts, it can serve as an alternative to Shannon entropy for quantifying the classification uncertainty when computing the pseudo-label weights [62]. It could also be used in domain adaptation contexts to improve test-time adaptation. For example, the class ambiguity could be used analogously to entropy regularization [63] for information maximization. This class ambiguity-based regularization could help models generalize better under domain shifts without additional training [64].

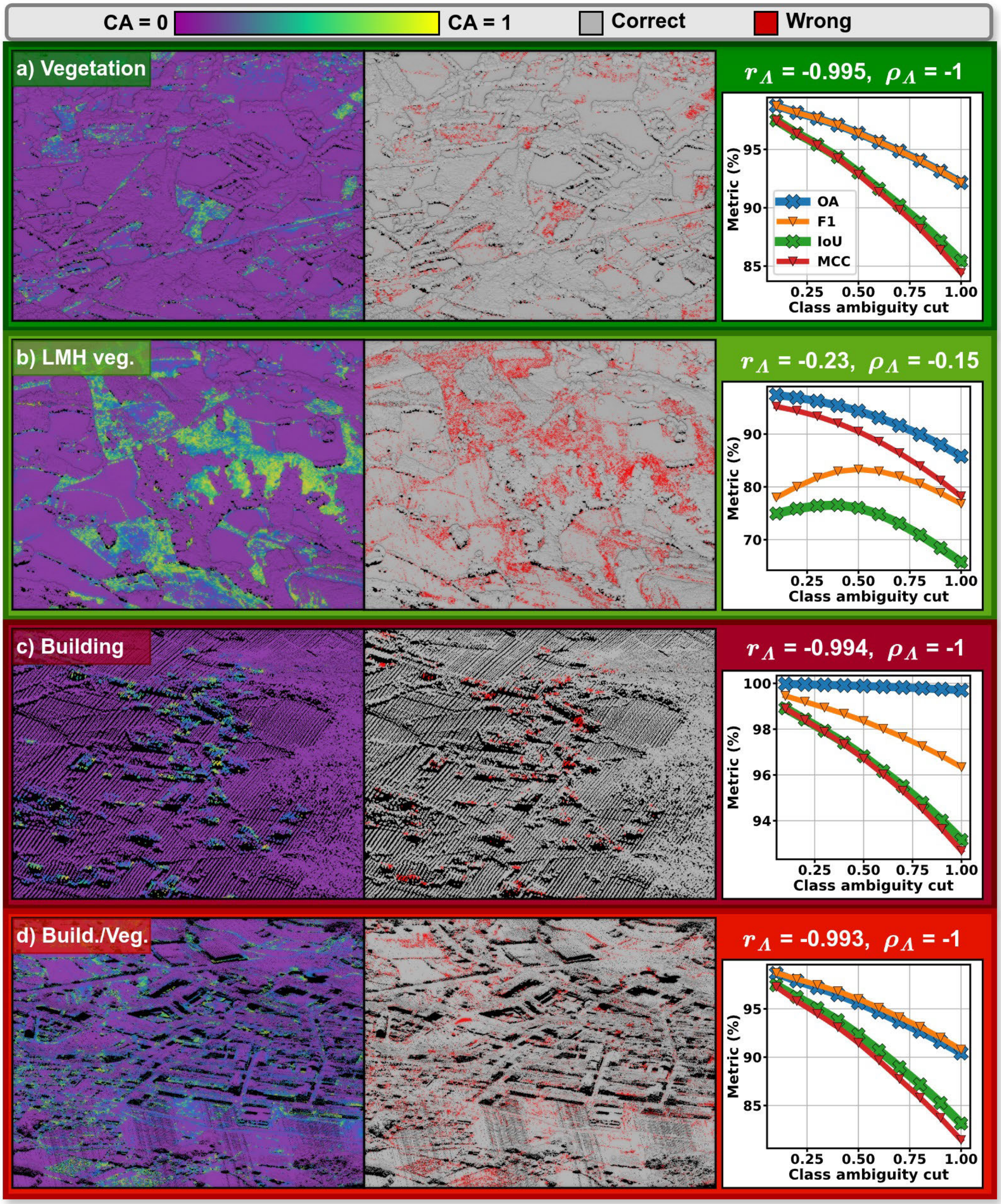


FIGURE 8. Uncertainty assessment for the four different classification tasks. The images on the left side represent the point-wise class ambiguity and the error mask in the point clouds. The figure on the right represents the evaluation and correlation metrics as a function of the class ambiguity cut. The selected metrics are the overall accuracy (OA), the F1-score (F1), the intersection over union (IoU), and Matthew's correlation coefficient (MCC). A class ambiguity cut of x implies that only points with an ambiguity less than or equal to x are considered to compute the metrics. Moreover, Pearson's correlation coefficient r_Λ and Spearman's correlation coefficient ρ_Λ between the class ambiguity cut and the F1 are shown.

C. FUTURE WORK

The most straightforward future challenge is to increase the scale of the case study. In this research, we have considered Galicia, which implies analyzing approximately $29,557\,\mathrm{km}^2$ of surface area. Extending the geographic area to a national scale would result in approximately $501,486\,\mathrm{km}^2$

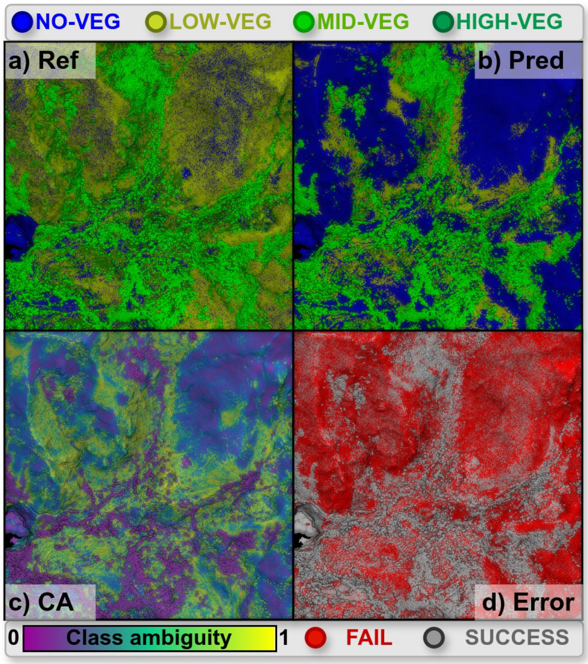


FIGURE 9. Uncertainty and error for low, mid, and high-vegetation classification using the SFL-Net model with (x, y, z), NIR, and RGB data. The labels are shown in a), the model's predictions in b), the class ambiguity in c), and the error mask in d). Most errors are concentrated in low-vegetation regions.

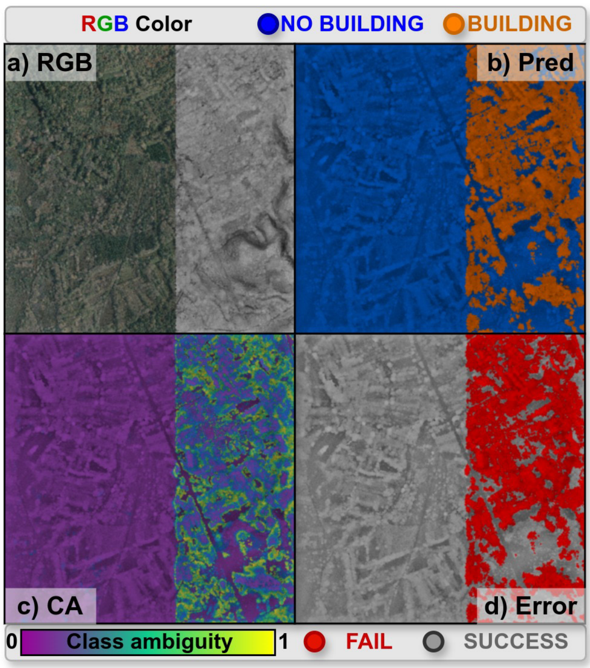


FIGURE 10. Uncertainty and error for building classification using the SFL-Net model with (x, y, z), NIR, and RGB data. The color is shown in a), the model's predictions in b), the class ambiguity in c), and the error mask in d). The errors are concentrated in the region that lacks RGB data.

($505,990\,\text{km}^2$ with 0.89% being water), using Spain as an example. Going even further, future works could aim at a continental scale, analyzing $4,101,308\,\text{km}^2$ if considering the European Union ($4,225,104\,\text{km}^2$ with 2.93% being water) or around ten million square kilometers if considering the whole European continent. Positive results will demonstrate the viability of laser-scanning-based remote sensing for understanding and monitoring the entire Earth's surface using highly accurate 3D data, thereby enabling large-scale industrial applications.

VI. CONCLUSION

This paper presents novel research using an ultra-large-scale dataset comprising approximately 36, 369 million points and a geographic extent of around $29,557\,\text{km}^2$. This notable dataset allows us to measure the generalization error of deep learning models for 3D point cloud classification more rigorously than is usually reported in the literature. Our experiments demonstrate that achieving accurate classifications (e.g., at least 95% of high-vegetation points correctly classified) on many tasks is possible even with 38% missing labels and high measurement error in the data.

Our second conclusion is that point-wise uncertainty metrics provide a reliable quantification that is strongly inversely correlated with correct classifications. Therefore, we propose class ambiguity as an alternative uncertainty metric that performs slightly better than entropy in the general case and is significantly more robust for problematic classifications, as indicated by Pearson's and Spearman's correlation coefficients. Reciprocally, we also expect that using class ambiguity in active learning contexts might help to minimize epistemic uncertainty more efficiently.

Finally, we closed the gap between regional and real-world airborne laser scanning (ALS) datasets in terms of geographic scale. This issue is fundamental for real-world applications because it is the most economical procedure for scanning large-scale regions, such as countries or continents. We show that deep learning can be successfully adapted to ultra-large-scale 3D point clouds, even when using low-resolution data between $0.5\,\text{pts/m}^2$ and $1.2\,\text{pts/m}^2$ with 20 cm vertical error and 30 cm planimetric error.

VII. SOFTWARE AND DATA AVAILABILITY

The source code and script files used to run the experiments are provided in a fork of the VirtuaLearn3D++ open-source scientific software at GitHub (https://github.com/albertoesmp/vl3d_galicia). The dataset for the experiments is publicly available on the web page of the Spanish Ministry of Transport and Sustainable Mobility (https://pnoa.ign.es/pnoa-lidar/segunda-cobertura).

ACKNOWLEDGMENT

The numerous experiments conducted on the FinisTerrae-III supercomputer were made possible thanks to the CESGA (Galician supercomputing center) and the open ALS dataset PNOA-II, provided by the Ministerio de Transporte y Movilidad Sostenible, Government of Spain.

REFERENCES

[1] J. Shan and C. Toth, *Topographic Laser Ranging and Scanning: Principles and Processing*, 2nd ed. Boca Raton, FL, USA: CRC Press, 2018.

[2] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.

[3] A. Nurunnabi, F. Teferle, A. Novo, J. Balado, and E. Ientilucci, "Derivation of tree stem curve and volume using point clouds," *Int. Arch. Photogramm., Remote Sens. Spatial Inf. Sci.-ISPRS Arch.*, vol. 48, pp. 81–88, 2024.

[4] T. Arumäe and M. Lang, "ALS-based wood volume models of forest stands and comparison with forest inventory data," *Forestry Stud.*, vol. 64, no. 1, pp. 5–16, Jun. 2016.

[5] H. Sun, Q. Ye, Q. Chen, L. Fu, Z. Xu, and C. Hu, "Tree canopy volume extraction fusing ALS and TLS based on improved PointNeXt," *Remote Sens.*, vol. 16, no. 14, p. 2641, Jul. 2024.

[6] H. Weiser, A. M. E. Pena, and B. Höfle, "How tree movement influences tree metrics derived from laser scanning point clouds," in *Proc. EGU Gen. Assem. Conf. Abstracts*, 2024, p. 1633.

[7] N. Fareed, J. P. Flores, and A. K. Das, "Analysis of UAS-LiDAR ground points classification in agricultural fields using traditional algorithms and PointCNN," *Remote Sens.*, vol. 15, no. 2, p. 483, Jan. 2023.

[8] R. Wang, D. Yang, Y. Ma, D. Liu, X. Wang, and H. Yang, "Research on crop fruit segmentation method based on point cloud," in *Proc. 9th Int. Conf. Virtual Reality (ICVR)*, May 2023, pp. 78–85.

[9] M. O. Mints, N. Theisen, P. Neubert, and D. Paulus, "Point-cloud-based change detection for steep slope vineyard agriculture," in *Proc. IEEE SENSORS*, Oct. 2023, pp. 1–4.

[10] S. Xiang and D. Li, "Research on plant growth tracking based on point cloud segmentation and registration," in *Proc. Int. Conf. Image Process., Comput. Vis. Mach. Learn. (ICICML)*, Oct. 2022, pp. 469–478.

[11] P. Orozco Carpio, M. Viñals, and M. López-González, "3D point cloud and GIS approach to assess street physical attributes," *Smart Cities*, vol. 7, no. 3, pp. 991–1006, Apr. 2024.

[12] G. Zhou and J. Guo, "Extraction of urban road surface from mobile laser scanning point clouds," in *Proc. 2nd Int. Conf. Algorithms, High Perform. Comput. Artif. Intell. (AHPCAI)*, Oct. 2022, pp. 798–801.

[13] A. M. Esmorís, D. L. Vilariño, D. F. Arango, F.-A. Varela-García, J. C. Cabaleiro, and F. F. Rivera, "Characterizing zebra crossing zones using LiDAR data," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 38, no. 13, pp. 1767–1788, Sep. 2023.

[14] J. Xiao, D. Oh, and K. Kang, *Hydrant Segmentation and Extraction Using Deep Learning Models for Large-Scale Urban Point Clouds*, 2024, pp. 289–298. [Online]. Available: https://doi.org/10.1061/9780784485262.030

[15] T. Jiang, Y. Wang, S. Liu, Q. Zhang, L. Zhao, and J. Sun, "Instance recognition of street trees from urban point clouds using a three-stage neural network," *ISPRS J. Photogramm. Remote Sens.*, vol. 199, pp. 305–334, May 2023.

[16] P. Wang, Y. Tang, Z. Liao, Y. Yan, L. Dai, S. Liu, and T. Jiang, "Road-side individual tree segmentation from urban MLS point clouds using metric learning," *Remote Sens.*, vol. 15, no. 8, p. 1992, Apr. 2023.

[17] X. Sun, B. Guo, C. Li, N. Sun, Y. Wang, and Y. Yao, "Semantic segmentation and roof reconstruction of urban buildings based on LiDAR point clouds," *ISPRS Int. J. Geo-Inf.*, vol. 13, no. 1, p. 19, Jan. 2024.

[18] J. Shao, W. Yao, P. Wang, Z. He, and L. Luo, "Urban GeoBIM construction by integrating semantic LiDAR point clouds with as-designed BIM models," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5701712.

[19] J. Zhao, H. Yu, X. Hua, X. Wang, J. Yang, J. Zhao, and A. Xu, "Semantic segmentation of point clouds of ancient buildings based on weak supervision," *Heritage Sci.*, vol. 12, no. 1, p. 232, Jul. 2024.

[20] S. Salamanca, P. Merchán, A. Espacio, E. Pérez, and M. J. Merchán, "Segmentation of 3D point clouds of heritage buildings using edge detection and supervoxel-based topology," *Sensors*, vol. 24, no. 13, p. 4390, Jul. 2024.

[21] J. Sobotka and M. Priesner, "Revitalization of the military shooting range using point clouds," in *Proc. Int. Conf. Mil. Technol. (ICMT)*, May 2023, pp. 1–5.

[22] R. J. Salgado-Garcia, N. Vila-Blanco, M. J. Carreira, and M. Nuñez-Garcia, "Efficient multi-modal whole heart segmentation via cascaded U-Net: A practical solution for clinical settings," in *Comprehensive Analysis and Computing of Real-World Medical Images*, X. Zhuang, W. Ding, F. Wu, S. Gao, L. Li, and S. Wang, Eds., Cham, Switzerland: Springer, 2025, pp. 158–167.

[23] L. Vaquero, Y. Xu, X. Alameda-Pineda, V. M. Brea, and M. Mucientes, "Lost and found: Overcoming detector failures in online multi-object tracking," in *Computer Vision–(ECCV)*, A. Leonardis, E. Ricci, S. Roth, O. Russakovsky, T. Sattler, and G. Varol, Eds., Cham, Switzerland: Springer, 2025, pp. 448–466.

[24] K. M. Hasib, M. R. Islam, S. Sakib, M. A. Akbar, I. Razzak, and M. S. Alam, "Depression detection from social networks data based on machine learning and deep learning techniques: An interrogative survey," *IEEE Trans. Computat. Social Syst.*, vol. 10, no. 4, pp. 1568–1586, Aug. 2023.

[25] K. Md Hasib, M. F. Mridha, M. Humaion Kabir Mehedi, K. Omar Faruk, R. Khatun Muna, S. Iqbal, M. R. Islam, and Y. Watanobe, "DCNN: Deep convolutional neural network with XAI for efficient detection of specific language impairment in children," *IEEE Access*, vol. 12, pp. 101660–101678, 2024.

[26] A. M. Esmorís, H. Weiser, L. Winiwarter, J. C. Cabaleiro, and B. Höfle, "Deep learning with simulated laser scanning data for 3D point cloud classification," *ISPRS J. Photogramm. Remote Sens.*, vol. 215, pp. 192–213, Sep. 2024.

[27] F. E. Fassnacht, J. C. White, M. A. Wulder, and E. Næsset, "Remote sensing in forestry: Current challenges, considerations and directions," *Forestry: Int. J. Forest Res.*, vol. 97, no. 1, pp. 11–37, Jan. 2024.

[28] W. Han, R. Wang, D. Huang, and C. Xu, "Large-scale ALS data semantic classification integrating location-context-semantics cues by higher-order CRF," *Sensors*, vol. 20, no. 6, p. 1700, Mar. 2020.

[29] Z. Zhang, R. Liu, E. Xie, and G. Zhang, "Large scale point cloud classification base on graph-MLP++," in *Proc. IEEE 9th Int. Conf. Data Sci. Adv. Anal. (DSAA)*, Oct. 2022, pp. 1–6.

[30] R. Liu, Z. Zhang, L. Dai, G. Zhang, and B. Sun, "MFTR-Net: A multi-level features network with targeted regularization for large-scale point cloud classification," *Sensors*, vol. 23, no. 8, p. 3869, Apr. 2023.

[31] W. Tan, N. Qin, L. Ma, Y. Li, J. Du, G. Cai, K. Yang, and J. Li, "Toronto-3D: A large-scale mobile LiDAR dataset for semantic segmentation of urban roadways," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Los Alamitos, CA, USA, Jun. 2020, pp. 797–806.

[32] J. Pan, K. Cao, B. Zhao, W. Li, and T. Zhang, "Semantic segmentation of large-scale point clouds by encoder–decoder shared MLPs with weighted focal loss," in *Proc. IEEE 21st Int. Conf. Cognit. Informat. Cognit. Comput. (ICCI*CC)*, Dec. 2022, pp. 153–159.

[33] Z. Zeng, Y. Xu, Z. Xie, W. Tang, J. Wan, and W. Wu, "Large-scale point cloud semantic segmentation via local perception and global descriptor vector," *Expert Syst. Appl.*, vol. 246, Jul. 2024, Art. no. 123269.

[34] S. M. Iman Zolanvari, S. Ruano, A. Rana, A. Cummins, R. Eduardo da Silva, M. Rahbar, and A. Smolic, "DublinCity: Annotated LiDAR point cloud and its applications," 2019, *arXiv:1909.03613*.

[35] S. Singh and J. Sreevalsan-Nair, "A distributed system for multiscale feature extraction and semantic classification of large-scale LiDAR point clouds," in *Proc. IEEE India Geosci. Remote Sens. Symp. (InGARSS)*, Dec. 2020, pp. 74–77.

[36] M. Turgeon-Pelchat, S. Foucher, and Y. Bouroubi, "Deep learning-based classification of large-scale airborne LiDAR point cloud," *Can. J. Remote Sens.*, vol. 47, no. 3, pp. 381–395, May 2021.

[37] H. Wang and D. Wang, "3D semantic understanding of large-scale urban scenes from LiDAR point clouds," in *Proc. 9th Int. Conf. Virtual Reality (ICVR)*, May 2023, pp. 86–92.

[38] Y. Zhao, X. Ma, B. Hu, Q. Zhang, M. Ye, and G. Zhou, "A large-scale point cloud semantic segmentation network via local dual features and global correlations," *Comput. Graph.*, vol. 111, pp. 133–144, Apr. 2023.

[39] T. Hackel, "Large-scale machine learning for point cloud processing," Ph.D. dissertation, Inst. Geodesy Photogramm., ETH Zürich, Zürich, Switzerland, 2018. [Online]. Available: https://www.research-collection.ethz.ch/entities/publication/f96f00a3-60b1-49ab-a60a-7c21299b7206

[40] Y. Zhang, Z. Li, Y. Xie, Y. Qu, C. Li, and T. Mei, "Weakly supervised semantic segmentation for large-scale point cloud," in *Proc. AAAI Conf. Artif. Intell.*, 2021, vol. 35, no. 4, pp. 3421–3429.

[41] Y. Li, Q. Lin, Z. Zhang, L. Zhang, D. Chen, and F. Shuang, "MFNet: Multi-level feature extraction and fusion network for large-scale point cloud classification," *Remote Sens.*, vol. 14, no. 22, p. 5707, Nov. 2022.

[42] F. Zhang and X. Xia, "Cascaded contextual reasoning for large-scale point cloud semantic segmentation," *IEEE Access*, vol. 11, pp. 20755–20768, 2023.

[43] Y. Zhao, J. Zhang, J. Ma, and S. Xu, "Large-scale semantic scene understanding with cross-correction representation," *Remote Sens.*, vol. 14, no. 23, p. 6022, Nov. 2022.

[44] Z. J. Tang and T.-J. Cham, "MPT-Net: Mask point transformer network for large scale point cloud semantic segmentation," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Oct. 2022, pp. 10611–10618.

[45] C. Gaydon, M. Daab, and F. Roche, "FRACTAL: An ultra-large-scale aerial LiDAR dataset for 3D semantic segmentation of diverse landscapes," 2024, *arXiv:2405.04634*.

[46] (2017). *OpenStreetMap Contributors*. [Online]. Available: https://planet.osm.org and https://www.openstreetmap.org

[47] A. Lomba, J. Ferreiro da Costa, P. Ramil-Rego, and E. Corbelle-Rico, "Assessing the link between farming systems and biodiversity in agricultural landscapes: Insights from Galicia (Spain)," *J. Environ. Manage.*, vol. 317, Sep. 2022, Art. no. 115335.

[48] (2024). Instituto de Estudos do Territorio. *Mapas de Galicia*. rúa de San Lázaro, s/n, 15707, Santiago de Compostela—A Coruña. [Online]. Available: https://mapas.xunta.gal

[49] A. Gómez-Pazo, A. Pérez-Alberti, and A. Trenhaile, "High resolution mapping and analysis of shore platform morphology in Galicia, Northwestern Spain," *Mar. Geol.*, vol. 436, Jun. 2021, Art. no. 106471.

[50] A. Géron, *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow. Concepts, Tools, and Techniques to Build Intelligent Systems*, 2nd ed. Sebastopol, CA, USA: O'Reilly Media, 2019.

[51] M. Kölle, "Forming a hybrid intelligence system by combining active learning and paid crowdsourcing for semantic 3D point cloud segmentation," Ph.D. dissertation, Inst. Photogrammetry, Faculty Aerosp. Eng. Geodesy, Univ. Stuttgart, München, Germany, 2023. [Online]. Available: https://publikationen.badw.de/en/049489886

[52] D. Fernández-Arango, F.-A. Varela-García, and A. M. Esmorís, "Methodology for identifying optimal pedestrian paths in an urban environment: A case study of a school environment in a Coruña, Spain," *Smart Cities*, vol. 7, no. 3, pp. 1441–1461, Jun. 2024.

[53] R. Q. Charles, H. Su, M. Kaichun, and L. J. Guibas, "PointNet: Deep learning on point sets for 3D classification and segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 77–85.

[54] C. R. Qi, L. Yi, H. Su, and L. J. Guibas, "PointNet++: Deep hierarchical feature learning on point sets in a metric space," 2017, *arXiv:1706.02413*.

[55] H. Thomas, C. R. Qi, J.-E. Deschaud, B. Marcotegui, F. Goulette, and L. Guibas, "KPConv: Flexible and deformable convolution for point clouds," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 6410–6419.

[56] X. Li, Z. Zhang, Y. Li, M. Huang, and J. Zhang, "SFL-NET: Slight filter learning network for point cloud semantic segmentation," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5703914.

[57] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes. The Art of Scientific Computing*, 3rd ed. Cambridge, U.K.: Cambridge Univ. Press, 2023, ch. 10, pp. 515–520.

[58] M. Weinmann, S. Urban, S. Hinz, B. Jutzi, and C. Mallet, "Distinctive 2D and 3D features for automated large-scale scene analysis in urban areas," *Comput. Graph.*, vol. 49, pp. 47–57, Jun. 2015.

[59] T. Hackel, J. D. Wegner, and K. Schindler, "Fast semantic segmentation of 3D point clouds with strongly varying density," *ISPRS Ann. Photogramm., Remote Sens. Spatial Inf. Sci.*, vol. 3, pp. 177–184, Jun. 2016.

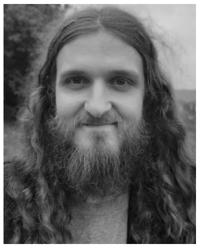
[60] H. Thomas, F. Goulette, J.-E. Deschaud, B. Marcotegui, and Y. LeGall, "Semantic classification of 3D point clouds with multiscale spherical neighborhoods," in *Proc. Int. Conf. 3D Vis. (3DV)*, Sep. 2018, pp. 390–398.

[61] J. Kirkpatrick, R. Pascanu, N. C. Rabinowitz, J. Veness, G. Desjardins, A. A. Rusu, K. Milan, J. Quan, T. Ramalho, A. Grabska-Barwinska, D. Hassabis, C. Clopath, D. Kumaran, and R. Hadsell, "Overcoming catastrophic forgetting in neural networks," *Proc. Nat. Acad. Sci. USA*, vol. 114, no. 13, pp. 3521–3526, 2017.

[62] P. Wang, W. Yao, and J. Shao, "One class one click: Quasi scene-level weakly supervised point cloud semantic segmentation with active learning," *ISPRS J. Photogramm. Remote Sens.*, vol. 204, pp. 89–104, Oct. 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0924271623002344

[63] P. Wang and W. Yao, "A new weakly supervised approach for ALS point cloud semantic segmentation," *ISPRS J. Photogramm. Remote Sens.*, vol. 188, pp. 237–254, Jun. 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0924271622001198

[64] P. Wang, W. Yao, J. Shao, and Z. He, "Test-time adaptation for geospatial point cloud semantic segmentation with distinct domain shifts," *ISPRS J. Photogramm. Remote Sens.*, vol. 229, pp. 422–435, Nov. 2025. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0924271625003338



ALBERTO M. ESMORÍS received the B.S. degree in computer science from Universitat Oberta de Catalunya, and the M.S. degree in high performance computing and the Ph.D. degree in computer science from Universidade de Santiago de Compostela, where he is currently pursuing the Ph.D. degree in mathematics.

He is researching for the Department of Mathematics, Universidade de Santiago de Compostela, and the Galician Research and Technology Center on Mathematics (CITMAga), and collaborating with the Computer Architecture Research Group, Universidade de Santiago de Compostela, and the 3DGeo Research Group, Universität Heidelberg. His research interests include high-performance computing, artificial intelligence, computer vision, computational algebraic geometry, applied algebra, and laser scanning simulation, with a special emphasis on 3D point clouds.

MIGUEL YERMO received the B.S. degree in physics and the M.S. degree in applied mathematics, in 2015 and 2017, respectively, and the Ph.D. degree in high-performance computing from Universidade de Santiago de Compostela, Spain, in 2024. He is currently an Assistant Professor with Universidade de Santiago de Compostela. His research career focuses on designing, implementing, and optimizing algorithms and data structures for processing large, irregular datasets in shared and distributed memory. He is researching and developing deep learning models for 3D point cloud segmentation.

SILVIA R. ALCARAZ (Member, IEEE) received the B.S. degree in computer science and the M.S. degree in high-performance computing from Universidade de Santiago de Compostela, Spain, in 2019 and 2020, respectively. She is currently pursuing the Ph.D. degree with the Research Center on Intelligent Technologies (CiTIUS), Universidade de Santiago de Compostela. Her main research interests include parallel programming, heterogeneous computing with a focus on distributed systems, performance portability, and high-level synthesis.

SAMUEL SOUTULLO received the B.S. degree in computer science from Universidade of Santiago de Compostela, Spain, in 2018, and the M.S. degree in game design and technology from the University of Gothenburg, in 2020. He is currently pursuing the Ph.D. degree with the Research Center on Intelligent Technologies (CiTIUS), Universidade de Santiago de Compostela. His main research interests include LiDAR data processing, intrinsic sensor parameter estimation, point cloud reconstruction, range image generation, geometric calibration, and high-performance computing for large-scale 3D perception applications.

FRANCISCO F. RIVERA received the B.S. degree in physics from Universidade de Santiago de Compostela, Spain, in 1986.

He is currently a Full Professor with Universidade de Santiago de Compostela. Throughout his career, he has supervised research and published extensively in the areas of computer-based applications, parallel processing, and computer architecture. His current research interests include the compilation of irregular codes for parallel and distributed systems, the analysis and prediction of performance on parallel systems, the design of parallel algorithms, memory hierarchy optimizations, GIS, image processing, and 3D point cloud computing.