MULTI-MODAL SEMANTIC MESH SEGMENTATION IN URBAN SCENES

Dominik Laupheimer*, Norbert Haala

Institute for Photogrammetry, University of Stuttgart, Germany
(dominik.laupheimer, norbert.haala)@ifp.uni-stuttgart.de

Commission II, WG II/6

KEY WORDS: Urban Scene Understanding, Semantic Segmentation, Multi-Modality, Textured Mesh, Point Cloud

ABSTRACT:

The semantic segmentation of the huge amount of acquired 3D data has become an important task in recent years. Meshes have evolved into a standard representation next to Point Clouds (PCs) - not least because of their great visualization possibilities. Compared to PCs, meshes have commonly smaller memory footprints while jointly providing geometrical and high-resolution textural information. For this reason, we opt for semantic mesh segmentation, which is a widely overlooked topic in photogrammetry and remote sensing yet. In this work, we perform an extensive ablation study on multi-modal handcrafted features adapting the Point Cloud Mesh Association (PCMA) (Laupheimer et al., 2020) which establishes explicit connections between faces and points. The multi-modal connections are used in a two-fold manner: (i) to extend per-face descriptors with features engineered on the PC and (ii) to annotate meshes semi-automatically by propagating the manually assigned labels from the PCs. In this way, we derive annotated meshes from the ISPRS benchmark data sets Vaihingen 3D (V3D) and Hessigheim 3D (H3D). To demonstrate the effectiveness of the multi-modal approach, we use well-established and fast Random Forest (RF) models deploying various feature vector compositions and analyze their performances for semantic mesh segmentation. The feature vector compositions consider features derived from the mesh, the PC or both. The results indicate that the combination of radiometric and geometric features outperforms feature sets of a single feature type only. Besides, we observe that relative height is the most crucial feature. The main finding is that the multi-modal feature vector integrates the complementary strengths of the underlying modalities. Whereas the mesh provides outstanding textural information, the dense PCs are superior in geometry. The multi-modal feature descriptor achieves the best performance on both data sets. It significantly outperforms feature sets that incorporate only features derived from the mesh by +7.37 pp and +2.38 pp for $mF1$ and Overall Accuracy (OA) on V3D. The registered improvement is +9.23 pp and +4.33 pp for $mF1$ and OA on H3D.

1. INTRODUCTION

3D data acquisition and processing have increasingly become feasible and important in the domain of photogrammetry and remote sensing in the past decade. Common representations are the modalities Point Cloud (PC) and mesh.

PC processing and interpretation are currently one of the most popular topics. PCs are unordered sets of points directly measured with Airborne Laser Scanning (ALS) or derived from images via Multi-View Stereo (MVS). In contrast, surface meshes are graphs consisting of vertices, edges, and faces that provide explicit adjacency information. The mesh is adaptive to the underlying geometry due to the non-uniformity and non-regularity of faces. This means planar surfaces are represented by a few large faces, whereas vivid areas are reconstructed by many small surface elements. Generally, the adaptiveness results in a less memory-consuming 3D representation compared to a PC. Another strength of meshes is the high-resolution texturing generating a realistic-looking 3D representation of the real world. We are aware that meshing of PCs is a non-trivial task. However, in our opinion, the mesh may replace unstructured PCs as the final user product for urban scenes in the future.

For these reasons, we strive for semantic segmentation of textured urban meshes. With this work, we want to account for the current hybridization trend and aim at semantic segmentation integrating information from PCs and meshes (hybrid semantics). Joint photogrammetric and LiDAR acquisition (hybrid acquisition) is state of the art for airborne systems and emerges for Un-

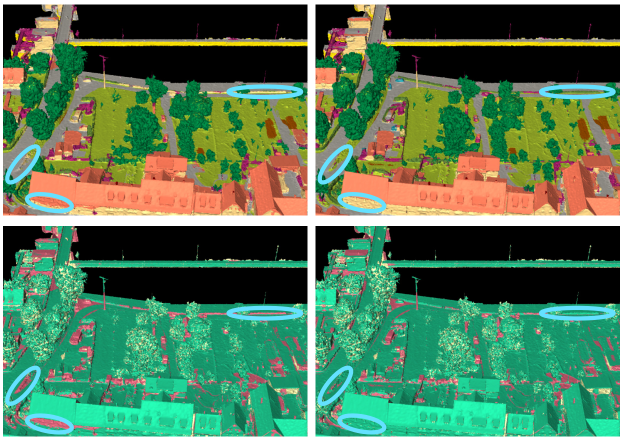

Figure 1. The *top* shows prediction results achieved with feature vector compositions $\mathcal{FS}_d$ (*left*, mesh features only), and $\mathcal{FS}_j$ (*right*, multi-modal features) (cf. Table 1) on the mesh in the lock area of Hessigheim 3D (H3D). The official class color code of the benchmark has been used throughout the paper (Kölle et al., 2021). The *bottom* shows the respective red-green plots indicating correct (*green*) and false (*red*) predictions. Faces with unknown ground truth are marked in *yellow*. The most obvious differences between the predictions are circled in *skyblue*.

[*] Corresponding author.

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-V-2-2022-267-2022

manned Airborne Vehicle (UAV)-based systems (Mandlburger et al., 2017; Cramer et al., 2018). Recently, Glira et al. (2019) proposed the hybrid orientation of complementary ALS PCs and aerial imagery. Haala et al. (2020) proved its potential for UAV data. The integrative character of the mesh facilitates data fusion "out-of-the-box" by utilizing LiDAR points and MVS points for the geometric reconstruction while leveraging high-resolution imagery for texturing (hybrid data storage). However, the information regarding the source modality is not encoded in the mesh vertices and hence, cannot be accessed in further processing steps, e.g., semantic segmentation.

Therefore, we make use of the Point Cloud Mesh Association (PCMA) presented in (Laupheimer et al., 2020) to recover the explicit one-to-many relationship between a face and points. The association mechanism allows propagating information between PCs and meshes in a subsequent information transfer. The scope of this work is not to design new distinguishing features. Instead, we utilize standard features that are commonly used for (semantic) segmentation tasks (cf. Section 2). Section 4.3 lists the features derived from the mesh or the PC. We briefly review the method and its improved implementation in Section 4.1.

We transfer information encoded in ALS PCs to meshes and enhance per-face descriptors to multi-modal feature vectors. We construct various descriptors considering features of different modalities, scales, and types (radiometry and geometry). The main contribution of this work is an extensive ablation study on the ISPRS benchmark data sets Vaihingen 3D (V3D) and Hessigheim 3D (H3D) (Niemeyer et al., 2014; Kölle et al., 2021) to investigate the impact of the deployed features. In particular, we analyze the benefit of the additional PC features by comparing the classifier performance with/without LiDAR support on the face level. We establish a feature-based pipeline for semantic mesh segmentation with well-established and fast RF models (Breiman, 2001). Figure 1 shows exemplary results for two dedicated feature vector compositions. In Section 5, we report the performance metrics for the RF models deploying various feature sets. We briefly present the data preparation and meshing in Section 3. The Ground Truth (GT) generation is discussed in Section 4.2.

2. RELATED WORK

The geospatial community put a big effort into the semantic segmentation of large-scale (LiDAR) PCs over the past few years. The recent rise of annotated urban mesh data may unleash the potential of mesh interpretation. Section 2.1 reviews important works for automated 3D scene interpretation - both for PCs and meshes. Machine Learning (ML) classifiers, particularly Deep Learning (DL) methods, rely on a huge amount of annotated data, wherefore GT generation has become a crucial task. We briefly review available GT for PCs and meshes in Section 2.2.

2.1 Semantic Segmentation of 3D Data

The 3D modalities PC and mesh facilitate expressive feature engineering as the underlying geometry can be accessed directly. A wide variety of features proved to be well-performing for semantic segmentation of ALS PCs employing RFs or support vector machines (Weinmann et al., 2015; Mallet et al., 2011; Chehata et al., 2009). The used features describe the local geometry or refine measured sensor data. Several works have revealed that the height above ground is the most expressive feature (Chehata et al., 2009; Guo et al., 2011; Kölle et al., 2019). Similarly, geometric contextual features can also be derived on the mesh (Kalogerakis et al., 2010). Rouhani et al. (2017) derive geometric and photometric features from a photogrammetric mesh, gather faces into so-called superfacets, and train a RF.

Although expressive feature calculation is feasible in 3D space, the rise of DL impacts semantic segmentation of 3D data. Griffiths and Boehm (2019) review the current state-of-the-art DL architectures for processing unstructured PCs. PointNet and its hierarchical, multi-scale successor PointNet++ (Qi et al., 2017a,b) constitute a milestone in DL-driven semantic PC segmentation operating directly on unstructured 3D PCs for the first time. KPConv introduces a continuous convolution defined by kernel points in the Euclidean space whose locations can be learned by the network (Thomas et al., 2019).

The emerging field of geometric DL extends basic DL operations to non-Euclidean domains such as graphs and manifolds (Bronstein et al., 2016). However, the encoded topological information complicates the application of DL and increases the computational burden during the representation learning. The majority of DL-based approaches proved their capacity on non-geospatial data sets (Qiao et al., 2019; Chang et al., 2018; Hu et al., 2021a). Knott and Groenendijk (2021) are the first who successfully adopted a well-performing DL approach from the computer vision community. They achieved to apply MeshCNN (Hanocka et al., 2019) to a real-world mesh at the expense of a reduced class catalog and restructuring the spatial mesh tiling to make use of local vicinity information.

Tutzauer et al. (2019) represent the mesh by its Centers of Gravity (COGs). The COG cloud reduces the complexity and simplifies further processing. Furthermore, it is robust against non-manifolds occurring in automatically generated meshes. The COG cloud differs from a common PC as it benefits from inherent mesh properties like the availability of high-resolution texture and adjacency knowledge. They adopted and enhanced the multi-branch 1D Convolutional Neural Network (George et al., 2017), which is a mixture of feature-engineering and feature-learning, and applied it to a 2.5D mesh.

In the light of the recent hybridization, multi-modal processing is expected to improve automated scene interpretation. Some works achieve multi-modality by rendering 2D views of the 3D scene, learning the semantic segmentation in image space, and finally, back-projecting the segmented 2D images onto the 3D modality. The detour via image space leverages well-performing and effective semantic image segmentation methods but comes along with information loss due to discretization, occlusions, and projection (Boulch et al., 2017; Lawin et al., 2017; He and Upcroft, 2013; Su et al., 2015; Kalogerakis et al., 2017). Therefore, other approaches augment 3D entities with features derived from high-resolution images instead of simply back-projecting per-pixel predictions to 3D space. Features are learned from images, projected to 3D, aggregated per 3D entity, and fed into a 3D network along with 3D geometry (Dai and Nießner, 2018; Jaritz et al., 2019). These approaches integrate 2D and 3D features in 3D space in an end-to-end-learning manner. Hu et al. (2021b) enhance the joint 2D-3D learning to joint 2D-3D semantic segmentation proposing the bidirectional projection network that simultaneously generates 2D and 3D predictions. The core of their symmetric network are two dedicated encoder-decoder networks for 2D and 3D coupled by the intermediate levels of the respective decoders.

The 2D-3D methods are characterized by integrating 3D data and imagery. In contrast, our approach explicitly links the entities of mesh and PC enabling flexible information sharing between the 3D modalities while incorporating image data implicitly via the mesh texture. To this end, we achieve multi-modal semantic segmentation of PCs and meshes with arbitrary

https://doi.org/10.5194/isprs-annals-V-2-2022-267-2022 | © Author(s) 2022. CC BY 4.0 License.

ML classifiers considering features from both representations. In (Laupheimer and Haala, 2021), we extended the inter-modal linking by explicitly associating the 3D modalities with imagery.

2.2 Ground Truth Availability

The vast majority of annotated geospatial data are LiDAR PCs captured with ALS or terrestrial laser scanning (Niemeyer et al., 2014; Hackel et al., 2017; Wichmann et al., 2018; Zolanvari et al., 2019). Recently, two annotated meshes have been provided to the community to foster semantic mesh segmentation. SUM provides a large-scale photogrammetric mesh covering the city of Helsinki (Gao et al., 2021). The mesh has been labeled semi-automatically by human-verified predictions achieved with an iteratively trained classifier. Likewise, H3D meshes have been annotated semi-automatically by propagating manual annotations from the PC to the respective mesh deploying PCMA (Kölle et al., 2021). In comparison to the pure photogrammetric mesh of Helsinki, H3D provides several epochs consisting of multi-modal meshes constructed from imagery and LiDAR data. To the best of our knowledge, H3D is the only data set that provides annotated multi-temporal PCs and meshes.

The PCMA-guided semi-automatic annotation reduces the tedious and time-consuming label work, specifically for large-scale data. PCMA can derive labeled meshes from any publicly available annotated PCs and vice versa, boosting the annotation of various representations. Ramirez et al. (2019) present a virtual reality tool that gamifies the manual labeling of meshes and PCs. Meshes have been a default data representation in the domain of computer vision for decades. Therefore, the development of semantic mesh segmentation is mainly driven by that community (cf. Section 2.1). However, they typically deal with meshes covering indoor scenes and single objects (Armeni et al., 2017; Shilane et al., 2004).

3. DATA PREPARATION

We utilize the manually annotated PCs of the ISPRS benchmark data sets V3D and H3D (Niemeyer et al., 2014; Kölle et al., 2021) to generate annotated meshes in a semi-automatically manner (cf. Section 4.2). V3D is typical for airborne large-scale country-wide mapping with moderate Ground Sampling Distance (GSD) of some centimeters (GSD = 8 cm) and a considerable time shift between ALS and nadir image data collection. As being captured in 2008, the ALS data features a point density of 4-8 $\frac{\text{points}}{\text{m}^2}$. In contrast, H3D provides high-resolution data with mainly synchronous data capture from a hybrid UAV sensor system and is representative of data collection at small-scale complex built-up areas (GSD = 2.5 cm and ALS PC density = 400-800 $\frac{\text{points}}{\text{m}^2}$). For both data sets, we generate textured and tiled meshes with `SURE 4.0.2` from nFrames (Rothermel et al., 2012). For H3D, we derive a hybrid mesh by fusing the simultaneously acquired oblique imagery and ALS data. Oblique images ensure proper texturing of vertical faces such as facades. We generate a purely photogrammetric mesh for V3D since the time-shift of imagery and ALS data is roughly one month.

4. METHODOLOGY

We transform the mesh into the COG cloud and attach a 1D feature vector to each COG. The linking of PCs and meshes as described in Section 4.1 enables the multi-modal enhancement of per-face feature vectors with features derived from the PC. The used features are described in Section 4.3. We annotate the generated meshes by propagating labels from the manually annotated PCs (cf. Section 4.2). Hence, ML classifiers can be trained to perform semantic mesh segmentation deploying arbitrary feature vector compositions (cf. Section 4.4).

4.1 Point Cloud Mesh Association (PCMA)

The PCMA explicitly links faces and points in a face-centered geometry-driven approach by establishing the one-to-many relationships between faces of the mesh and points of the PC (cf. Figure 2). We briefly describe the key steps of the association method for the understanding and refer the interested reader to (Laupheimer and Haala, 2021) for more details.

Each face f (represented by its COG) is assigned with n_{pts} points that represent the same surface by executing the following three steps for each f: (i) clipping of the point cloud $\mathcal{P}$ to a spherical vicinity of f ($\mathcal{P} \rightarrow \mathcal{P}'_f$), (ii) filtering of *out-of-face* points ($\mathcal{P}'_f \rightarrow \mathcal{P}''_f$), and (iii) filtering of *off-the-face* points ($\mathcal{P}''_f \rightarrow \mathcal{P}'''_f$). The remaining subset $\mathcal{P}'''_f$ is the set of n_{pts} points linked to f. *Out-of-face points* are points that are not enclosed by the face borders when projected orthogonally onto the face plane. *Off-the-face points* do not coincide with the face plane, i.e., they are below or above the face surface. A manually set threshold θ determines whether a point coincides with a face or not. Both point types are not mutually exclusive and exist due to discrepancies between the surface mesh and the ALS PC featuring multi-target capability. Geometric simplifications and geometry differences (e.g., due to 2.5D mesh geometry or asynchronous data acquisition) increase the structural differences.

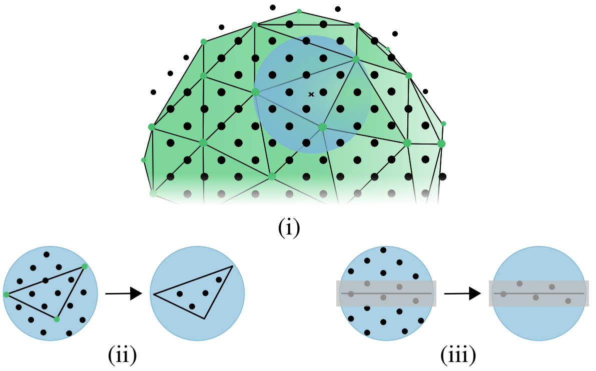

Figure 2. Sequential association steps (i) - (iii) of the PCMA to filter point cloud $\mathcal{P}$ for each face f: $\mathcal{P} \rightarrow \mathcal{P}'_f \rightarrow \mathcal{P}''_f \rightarrow \mathcal{P}'''_f$. *(i)*: Clipping of the PC (*black dots*) to the spherical vicinity (*blue*) of the considered face. Its COG is marked with a *black cross*. The mesh surface and its vertices are depicted in *green*. *(ii)*: Filtering of *out-of-face points* based on the clipping result (orthogonal view concerning the triangular face surface depicted by its edges marked in *black*). *(iii)*: Filtering of *off-the-face points* (side view with respect to the triangular face). The face is depicted as *black* line. The threshold band is marked in *gray*.

Concerning the scalability of the multi-modal association to the huge amount of acquired real-world data, we process data in a tile-wise fashion reducing the hardware requirements. Therefore, we impose the given mesh tiling on the PC and execute the previously described association steps for each tile. The parallel processing of tiles speeds up the association.

Since we have to compensate several structural discrepancies between PCs and 2.5D/3D meshes when processing real-world data, we use a more sophisticated adaptive thresholding with an arbitrary user-defined number of filter levels (Laupheimer and

Haala, 2021). Laupheimer et al. (2020) discuss in detail the particular challenges for the PCMA owing to 2.5D meshes and jointly annotate the respective 3D PC. The challenges and limitations of the association mechanism along with storage handling are discussed by Laupheimer and Haala (2021).

The established links between the entities of the two modalities enable the information transfer between mesh and PC. As we strive for semantic mesh segmentation, we transfer information from the PC to the mesh. The many-to-one relationship calls for information aggregation on the face level. For each face, we derive robust median features from the linked subset $\mathcal{P}_f'''$ (cf. Section 4.3). The propagated attributes may embrace sensor-intrinsic and handcrafted features, such as pulse characteristics and sophisticated engineered quantities. Analogously, majority votes determine the per-face labels as transferred from the associated points (cf. Section 4.2). The aggregated features and labels are attached to the COG cloud. Non-associated faces are marked with -1 labels and receive zeroed median features (Laupheimer et al., 2020).

Figure 3 shows the PC and the respective 3D mesh colored by the labels, the number of returns, and the reflectance. The annotations and visualized features have been automatically transferred from the manually annotated PC deploying PCMA-steered information propagation. Obviously, there are almost no differences notable between the modalities. This shows that the information transfer works reasonably - both for discrete and continuous quantities. Moreover, the filtering effect of the median aggregation becomes visible for the transferred features. For instance, roofs appear to be more monotonous on the mesh than on the PC - particularly for the colorized reflectance.

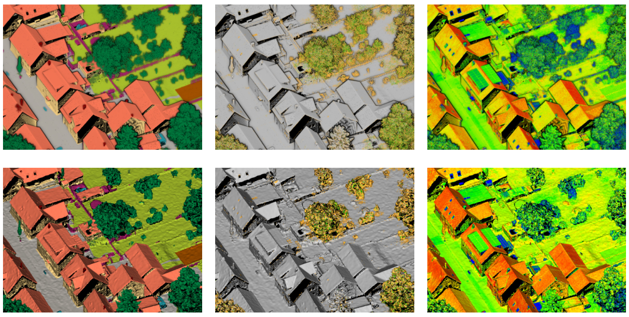
Figure 3. The *top row* shows the PC of H3D colored by manually assigned labels (*left*), number of returns (*center*), and reflectance values (*right*). The *bottom row* shows the counterparts on the mesh where all quantities have been automatically transferred from the PC deploying PCMA. Faces that cannot be linked with points are depicted in *black*. Please note the colorization on the two modalities does not match perfectly due to normal-depending shading, which is necessary to uncover structure, i.e., to generate an illusion of depth.

4.2 Ground Truth Generation

As discussed in Section 2.2, data annotation is a highly relevant task. The multi-modal entity linking allows the sharing of manually attached labels across modalities and boosts the annotation process by limiting the time-consuming and expensive labeling effort to a single representation. To this end, we utilize the manually annotated PC data of V3D and H3D to propagate their labels to the respective meshes leveraging the established entity connections. The per-face label is determined via majority voting. Please note, the semi-automatically annotated H3D mesh data is provided to the community as part of the H3D benchmark data (Kölle et al., 2021).

Figure 3 (*left*) showcases the semi-automatically derived mesh annotation by the example of H3D. The figure qualitatively verifies the effectiveness of the explicit entity linking. Additionally, we cross-checked the propagated labels by a small manually labeled subset of the H3D mesh. 96.7 % of the transferred labels match with the manually provided labels on the mesh. For comparison, a simple Nearest Neighbor (NN) interpolation between the modalities correctly annotates 86.8 % of the surface. Figure 4 contrasts the PCMA-steered propagation to the NN interpolation. Our propagation method transfers information only where an explicit connection is given. The provided annotations are consistent with the source modality as PCMA-steered transfer does not introduce unreliable information in data gaps like the interpolation, which increases label noise and feature noise. The *last column* depicts that our approach manages to deal with complex structures, e.g., fine structures in the lock area.

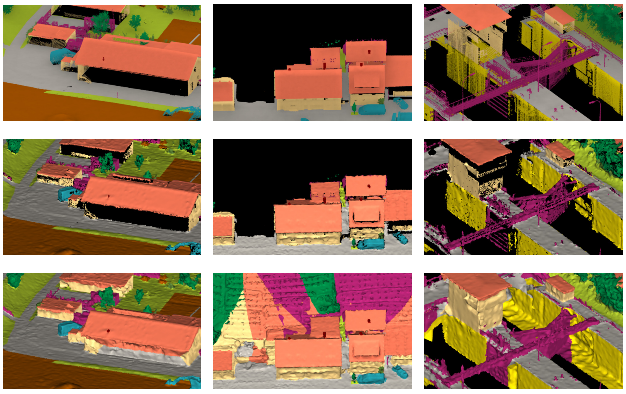
Figure 4. Effect of different propagation methods. The annotations of the manually labeled H3D PC (*top*) are propageted to the mesh by a) PCMA-steered transfer (*center*) or b) nearest neighbor interpolation (*bottom*).

4.3 Multi-Modal Feature Engineering

Handcrafted features are categorized according to their a) type (geometric and radiometric features), b) scale (per-entity and contextual features), and c) modality (derived from PC and mesh). Whereas per-entity features consider a single entity, contextual features gather information from adjacent instances. Thereby, contextual features implicitly act as spatial regularization causing spatially smooth labeling (Schindler, 2012).

The COG cloud allows PC-like handling of the mesh while preserving features that have been calculated on the mesh topology. For this reason, local vicinities can be defined identically on both modalities. We define multiple spherical neighborhoods $\mathcal{NH}_i^r$ with radius/scale $r \in \{1\,\mathrm{m}, 2\,\mathrm{m}, 3\,\mathrm{m}, 5\,\mathrm{m}\}$ for each entity to establish multiple levels of abstraction. The set of radii is inspired by several works (Niemeyer et al., 2014; Blomley and Weinmann, 2017; Kölle et al., 2019). As a consequence, the feature regularization on the mesh neglects the mesh topology but efficiently derives multi-scale PC and mesh features. For instance, the computationally expensive texture patch extraction has to be done only for a single scale, i.e., the face level. We calculate the Gaussian weighted average within $\mathcal{NH}_i^r$ to generate the contextual versions of the per-entity features.

Apart from modality-related features, we robustly calculate the *relative height* for each entity by subtracting the terrain height from the entity height. The *relative height of face* describes the height above ground of the respective COG. The required terrain height is interpolated bilinearly from the respective Digital Terrain Model (DTM).

Point Cloud Features. In addition to features provided by the LiDAR sensor itself such as *reflectance* and *echo ratio*, we calculate multi-scale features that proved to be well-performing for semantic segmentation of ALS PCs (cf. Section 2).

To describe the local point distribution, we estimate the covariance matrix of coordinates (also known as structure tensor) for the local neighborhood of each point. The principal component analysis provides the eigenvalues and eigenvectors of the structure tensor. Eigenvalue-based features refine the eigenvalues to more comprehensive features: *linearity*, *planarity*, *anisotropy*, *sphericity*, *change of curvature*, *omnivariance*, *eigenentropy*, and *sum of eigenvalues* (Mallet et al., 2011; Weinmann et al., 2015). Computing eigenvalues and eigenvectors eventually means fitting a local plane. The eigenvector corresponding to the smallest eigenvalue represents the normal vector of that local plane. We extend the feature set by plane-based features *verticality*, *roughness*, and *inclination*. Furthermore, we extend the point descriptor with density characteristics of the local vicinity by *volume density* and *surface density* (Weinmann et al., 2013). The presented features are median-aggregated per associated face and attached to the COG cloud.

Mesh Features. We parse the textured meshes to derive geometric and radiometric features for each face and attach them to the COG cloud. Geometric features are computed by employing *PyMesh* (Zhou, 2018). We briefly introduce the calculated features in the following.

Features may be calculated per face or vertex. Per-face features such as *face normal*, *face area*, *relative height of face*, and radiometric features are straight-forward to calculate. The geometric per-vertex features inherently consider contextual information by exploiting the 1-ring neighborhood $\mathcal{NH}_i^{①}$of a vertex v_i. $\mathcal{NH}_i^{①}$comprises all adjacent vertices, incident edges, and incident faces of v_i. *Valence* defines the number of edges incident to a vertex and hence describes its degree of connectivity. The *dihedral angle* of a vertex represents the maximum dihedral angle between its incident faces and is a measure for surface flatness. A dihedral angle of a vertex is the angle between two faces sharing a common edge that incidents at the considered vertex. Similarly, the *Gaussian curvature* and *mean curvature* describe the surface topography.

To obtain radiometric features, we extract the texture patch T_f for each face f from the lightweight texture atlas. The predominant *face color* is captured by the median RGB tuple values of T_f. We use RGB color information, as well as its transformed pendant in HSV color space, to ensure lighting independent features. To increase color expressivity, we calculate the *face color variance* per face for both color spaces. Additionally, we calculate 8-binned histograms for each color channel to derive the *face color signature*. Binning of 8 was set heuristically to balance the number of empty bins and the required memory.

4.4 Setup for Feature Vector Composition Ablation Study

Leveraging the flexible feature-based COG representation in combination with PCMA-steered information propagation allows for an extensive ablation study with various feature vector compositions considering multi-scale, multi-modal, multi-type features. We perform the feature-driven ablation study on the two benchmark data sets V3D and H3D with significantly different properties. For each face, we generate various feature sets $\mathcal{FS}$ consisting of measured and engineered features derived from the mesh and the PC (cf. Section 4.3). Table 1 gives an overview of the studied $\mathcal{FS}$. Feature vector compositions $\mathcal{FS}_a - \mathcal{FS}_d$ use features that have been derived on the mesh ("mesh-only"). $\mathcal{FS}_e - \mathcal{FS}_i$ deploy features for each face that have been derived on the PC and propagated to the mesh ("PC-only"). Feature vector $\mathcal{FS}_j$ combines the entire set of "mesh-only" and "PC-only" features in a multi-modal descriptor for each face on the mesh.

We opt for well-established and fast RF models to investigate the impact of deployed features. The features are robustly standardized according to the central moments of the dedicated train sets. For comparability, the trained RF variants consistently utilize 100 trees with a depth of 18 nodes (empirically determined by grid search). All trained models deploy the same weighting strategy, which considers class-dependent and sample-dependent surface-aware weights. In particular, we set face areas as sample weights to account for the non-uniformity. To tackle the class imbalance, we weight classes inversely proportional to the class-specific covered area.

$\mathcal{FS}$		Description
"mesh-only"	a	Mesh-intrinsic geometric mesh features (i.e., without relative height)
	b	Geometric mesh features (i.e., mesh-intrinsic + relative height)
	c	Radiometric mesh features (i.e., texture)
	d	Geometric & radiometric mesh features (i.e., all mesh features)
"PC-only"	e	PC-intrinsic geometric PC features (i.e., without relative height)
	f	Geometric PC features (i.e., PC-intrinsic + relative height)
	g	Radiometric PC features (i.e., reflectance)
	h	Geometric & radiometric PC features
	i	Geometric & radiometric PC features and echo characteristics (i.e., all PC features)
multi-modal	j	Geometric & radiometric features from both modalities and echo characteristics (i.e., fusion of $\mathcal{FS}_d$ and $\mathcal{FS}_i$)

Table 1. Overview of the feature vector compositions $\mathcal{FS}$ sorted into groups "mesh-only", "PC-only" and "multi-modal". The computed features are described in Section 4.3.

5. ABLATION STUDY: ANALYSIS OF DIFFERENT FEATURE VECTOR COMPOSITIONS

The trained RF models are evaluated on the dedicated test sets of V3D and H3D with semi-automatically generated labels as described in Section 4.2. We discuss the versatile feature vector compositions $\mathcal{FS}$ listed in Table 1 by the achieved performance metrics of the respective RF models and summarize the main findings. Table 2 lists surface-weighted per-class $F1$-scores, $mF1$-scores and OAs for the deployed $\mathcal{FS}$ on both data sets. The COG predictions are weighted by the area of the respective face. $mF1$-score is the mean across per-class $F1$-scores. Figure 1 shows the prediction results for the semantic mesh segmentation deploying $\mathcal{FS}_d$ and $\mathcal{FS}_j$.

We use the python package `scikit-learn` to inspect the relevance of the deployed features. The relevance analysis reveals that multi-scale contextual features are more important than per-instance features for both radiometric and geometric features. The most relevant mesh-intrinsic features are *face normal* (more

Data	$\mathcal{FS}$	*F1-score* [%]								$mF1$	OA
		Low Veg.	*I. Surf.*	*Car*	*Fence/ Hedge*	*Roof*	*Facade*	*Shrub*	*Tree*	[%]	[%]
V3D	*a*	29.53	61.73	6.94	1.33	66.26	38.87	5.92	53.30	32.98	51.49
	b	53.09	70.17	14.34	3.96	85.23	38.48	33.62	62.52	45.18	65.97
	c	52.89	78.70	16.92	6.62	81.51	28.77	25.68	58.11	43.65	66.34
	d	75.90	91.19	28.87	10.07	93.16	53.18	37.86	77.89	58.51	82.17
	e	50.47	75.46	32.27	16.03	78.83	61.39	28.14	66.15	51.09	65.99
	f	60.31	77.42	33.94	17.58	88.20	61.93	40.02	67.96	55.92	72.25
	g	39.10	74.23	15.56	16.29	40.03	13.11	15.38	31.82	30.69	44.45
	h	77.34	93.17	63.85	16.78	87.93	61.55	41.03	67.88	63.69	79.65
	i	**77.51**	93.09	**63.95**	**18.37**	88.22	61.98	**41.19**	68.00	64.04	79.84
	j	77.39	**93.42**	55.38	13.57	**94.73**	**71.70**	39.88	**80.98**	**65.88**	**84.55**

Data	$\mathcal{FS}$	*F1-score* [%]											$mF1$	OA
		Low Veg.	*I. Surf.*	*Vehicle*	*U. Furn.*	*Roof*	*Facade*	*Shrub*	*Tree*	*Gravel/ Soil*	*V. Surf.*	*Chimney*	[%]	[%]
H3D	*a*	58.59	54.86	5.08	25.87	61.27	63.54	19.62	85.05	11.97	56.95	43.52	44.21	59.90
	b	69.99	59.67	11.27	29.50	85.60	67.78	25.75	88.26	11.44	65.85	56.79	51.99	68.67
	c	70.97	71.15	12.25	29.32	73.39	52.86	23.66	69.23	26.12	48.70	7.21	44.08	63.78
	d	86.32	85.98	15.51	46.32	86.54	68.78	45.26	90.84	27.06	67.04	49.81	60.86	80.51
	e	73.52	69.21	34.58	37.53	83.21	76.09	57.86	92.94	14.70	43.43	56.03	58.10	72.70
	f	76.40	71.47	38.89	41.21	90.70	78.84	59.06	93.35	19.93	54.08	69.72	63.06	76.35
	g	46.62	25.70	12.35	19.13	47.39	29.91	10.99	76.91	17.23	5.50	17.36	28.10	40.87
	h	76.96	71.94	51.45	42.64	90.13	78.47	60.54	93.53	29.42	48.59	**71.97**	65.06	77.03
	i	76.78	72.24	**52.26**	43.61	90.32	78.75	60.73	93.95	29.12	49.17	71.62	65.32	77.16
	j	**88.52**	**88.25**	45.94	**51.89**	**90.91**	**82.84**	**61.04**	**94.85**	**31.52**	**67.38**	67.81	**70.09**	**84.84**

Table 2. Semantic segmentation results on the test site of V3D (*top*) and H3D (*bottom*) achieved with RF models deploying various feature vector compositions $\mathcal{FS}$ (cf. Table 1). The performance metrics are weighted by the covered class area. Best performing metrics are marked in *bold*; worst are *underlined*.

precisely its third component) and the *face color*, underlining the strengths and key properties of the meshed representation. *Verticality* and *inclination* are important features derived from the PC. These features can be seen as the PC-pendants of the *face normal*. The most crucial feature is *relative height* δ_h. Categorically, the incorporation of δ_h improves the performance by decreasing the confusion between the majority of class pairs, e.g., *Shrub* vs. *Tree*. However, it increases the confusion for a few class pairs where δ_h is not a distinctive feature like for separating *Impervious Surface* from *Low Vegetation*. Its importance is reflected by the performance discrepancy between "mesh-only" configurations $\mathcal{FS}_a$ & $\mathcal{FS}_b$ and the "PC-only" compositions $\mathcal{FS}_e$ & $\mathcal{FS}_f$ respectively which both differ by δ_h only. Adding δ_h to modality-intrinsic geometric feature sets $\mathcal{FS}_a$ and $\mathcal{FS}_e$ improves the performance significantly for both $\mathcal{FS}$ groups. For instance, δ_h improves $mF1$-score and OA by 12.20 pp and 14.48 pp in "mesh-only" mode and by 4.83 pp and 6.26 pp in "PC-only" mode on V3D.

The δ_h-induced performance gain is significantly larger in "mesh-only" than in "PC-only" mode (factor 2.3–2.5 for V3D and factor 1.6–2.4 for H3D) hinting at the superiority of geometric features derived from the PC. Standard geometric PC-intrinsic features seem to be more expressive than default geometric mesh-intrinsic features regardless of the underlying data. Considering "PC-only" configurations $\mathcal{FS}_e - \mathcal{FS}_g$, we note that geometric feature vectors perform better than radiometric feature vectors for both data sets. The superiority of $\mathcal{FS}_e$ and $\mathcal{FS}_f$ over $\mathcal{FS}_g$ indicates the strong geometric information encoded in the PC features. For example, $\mathcal{FS}_f$ outperforms $\mathcal{FS}_g$ by 25.23 pp and 27.80 pp for $mF1$-score and OA respectively on V3D. Table 2 reveals that $\mathcal{FS}_g$ achieves worst or close-to-worst $F1$-scores for almost all classes for V3D and H3D. $\mathcal{FS}_g$ is the worst feature composition in total.

On the contrary, when considering the "mesh-only" configurations $\mathcal{FS}_a - \mathcal{FS}_c$, we notice that the radiometric feature sets perform better than mesh-intrinsic geometric feature sets for both data sets and can roughly compete with $\mathcal{FS}_b$. For V3D, $\mathcal{FS}_c$ significantly outperforms $\mathcal{FS}_a$ by more than 10 pp for both global performance metrics. The performance metrics of radiometric feature sets $\mathcal{FS}_c$ and $\mathcal{FS}_g$ show that texture information from the mesh is superior to the natively available radiometric information on the PC (i.e., reflectance). In contrast, the comparison of geometric "PC-only" configurations $\mathcal{FS}_e$ and $\mathcal{FS}_f$ with their "mesh-only" equivalents $\mathcal{FS}_a$ and $\mathcal{FS}_b$ show that geometric information derived from the PC significantly outperforms geometric information derived on the mesh. The findings validate our initial assumption: PCs are characterized by high-quality geometry, whereas meshes provide high-quality textural information. We are aware of the fact that these inter-modal comparisons are not entirely fair due to differing feature counts. The plentitude of geometric PC features uses 68 more features than the geometric $\mathcal{FS}$ of "mesh-only" compositions. Likewise, $\mathcal{FS}_c$ entails several textural features, whereas $\mathcal{FS}_g$ encompasses merely a handful of reflectance features. However, the features are derived straightforwardly on each modality and hence, implicitly encode and accentuate the strengths of both modalities. The feature design is steered inevitably by the modalities' properties and maps the balance of power between the two modalities.

Feature vectors that combine both feature types perform better at the global scale than configurations that incorporate only either geometry or radiometry for "mesh-only" and "PC-only" mode. Geometric features are util for the separation of radiometrically similar classes. Radiometric features help to separate geometrically similar classes. $\mathcal{FS}_d$ achieves the best performance for the "mesh-only" configurations on both data sets. Its "PC-only" counterpart $\mathcal{FS}_h$ achieves roughly on par performance like the best performing $\mathcal{FS}_i$ of "PC-only" configurations. $\mathcal{FS}_i$ enhances $\mathcal{FS}_h$ with echo characteristics and achieves marginally better performance metrics (≤ +0.35 pp for both global metrics). Comparing $\mathcal{FS}_h$ with $\mathcal{FS}_d$, we see that $\mathcal{FS}_h$ outperforms $\mathcal{FS}_d$ in terms of $mF1$-score for both data sets. Their comparison already documents the utility of the proposed multi-modal entity linking and the subsequent information transfer for semantic mesh segmentation although the compared descripors do not exploit the entire multi-modal potential.

$\mathcal{FS}_j$ integrates multi-scale geometric and radiometric features of both modalities outperforming other feature vector compositions in terms of $mF1$-scores and OA values. Table 2 reveals that $\mathcal{FS}_j$ achieves best $F1$-scores for almost all classes. Compared to the best "mesh-only" configuration (i.e., $\mathcal{FS}_d$), the multi-modal semantic segmentation achieves +7.37 pp and +2.38 pp for $mF1$ and OA on V3D. Similarly, $\mathcal{FS}_j$ outperforms $\mathcal{FS}_d$ by 9.23 pp and 4.33 pp for $mF1$ and OA on H3D. The significantly improved performance metrics demonstrate the superiority of multi-modal semantic mesh segmentation.

6. CONCLUSIONS AND OUTLOOK

In this paper, we evaluated the impact of multi-scale, multi-type, and multi-modal features on semantic mesh segmentation deploying RF models with various feature vector compositions. The integration of features derived from the PC was accomplished with PCMA, which explicitly links points and faces. We embedded PCMA in our semantic segmentation pipeline in two ways: (i) to enhance per-face feature descriptors with PC features and (ii) to generate annotated meshes of publicly available manually labeled PC data.

We run the performance analyses on the generated and semi-automatically annotated meshes of the ISPRS benchmark data sets V3D and H3D featuring considerably different point densities, GSDs and time-shifts between image and ALS data capture. The ablation study showed the efficacy and benefits of the PCMA for semantic mesh segmentation. Feature vectors deploying only features derived from the PC already outperform mesh-related feature sets. In particular, the enhancement of the face descriptors with features derived from the PC causes a significant performance boost. The multi-modal feature vector outperforms feature sets that incorporate only features derived from the mesh by +7.37 pp and +2.38 pp for $mF1$ and OA on V3D. The registered improvement is +9.23 pp and +4.33 pp for $mF1$ and OA on H3D. The multi-modal feature sets correctly predict 84.55–84.84 % of the surface area. The analysis revealed that the mesh provides superior textural information, while the dense PCs are superior in geometry. Hence, the multi-modal feature vector integrates the complementary strengths of underlying modalities. By these means, the multi-modality outperforms feature sets that combine radiometric and geometric features from a single modality only.

Regardless of the feature vector composition, *relative height* showed to be the most relevant feature.

PCMA facilitates the information transfer between PCs and meshes in both directions. Therefore, we would like to leverage the established entity connections to consistently segment the given LiDAR PCs with predictions from the 3D mesh in the future. So far, we have tested airborne scenarios with ALS PCs only. We plan to test the pipeline with terrestrial data or PCs derived from persistent scatterer interferometry. Moreover, we want to extend the information transfer to image space as proposed in Laupheimer and Haala (2021).

REFERENCES

Armeni, I., Sax, A., Zamir, A. R., Savarese, S., 2017. Joint 2D-3D-Semantic Data for Indoor Scene Understanding. *ArXiv*, abs/1702.01105.

Blomley, R., Weinmann, M., 2017. Using Multi-Scale Features for the 3D Semantic Labeling of Airborne Laser Scanning Data. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, IV-2/W4, 43–50.

Boulch, A., Saux, B. L., Audebert, N., 2017. Unstructured Point Cloud Semantic Labeling Using Deep Segmentation Networks. I. Pratikakis, F. Dupont, M. Ovsjanikov (eds), *Eurographics Workshop on 3D Object Retrieval*, The Eurographics Association.

Breiman, L., 2001. Random Forests. *Machine Learning*, 45(1), 5–32. https://doi.org/10.1023/A:1010933404324.

Bronstein, M. M., Bruna, J., LeCun, Y., Szlam, A., Vandergheynst, P., 2016. Geometric deep learning: going beyond Euclidean data. *CoRR*. http://arxiv.org/abs/1611.08097.

Chang, J., Gu, J., Wang, L., Meng, G., Xiang, S., Pan, C., 2018. Structure-aware convolutional neural networks. S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, R. Garnett (eds), *Advances in Neural Information Processing Systems 31*, Curran Associates, Inc., 11–20.

Chehata, N., Guo, L., Mallet, C., 2009. Airborne LIDAR feature selection for urban classification using random forests. *ISPRS Archives*, 38.

Cramer, M., Haala, N., Laupheimer, D., Mandlburger, G., Havel, P., 2018. Ultra-High Precision UAV-Based LiDAR and Dense Image Matching. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLII-1, 115–120.

Dai, A., Nießner, M., 2018. 3DMV: Joint 3D-Multi-View Prediction for 3D Semantic Scene Segmentation. *CoRR*, abs/1803.10409. http://arxiv.org/abs/1803.10409.

Gao, W., Nan, L., Boom, B., Ledoux, H., 2021. SUM: A Benchmark Dataset of Semantic Urban Meshes. *ISPRS Journal of Photogrammetry and Remote Sensing*, 179, 108-120.

George, D., Xie, X., Tam, G. K. L., 2017. 3D Mesh Segmentation via Multi-branch 1D Convolutional Neural Networks. *CoRR*, abs/1705.11050. http://arxiv.org/abs/1705.11050.

Glira, P., Pfeifer, N., Mandlburger, G., 2019. Hybrid Orientation of Airborne LiDAR Point Clouds and Aerial Images. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, IV-2/W5, 567–574.

Griffiths, D., Boehm, J., 2019. A Review on Deep Learning Techniques for 3D Sensed Data Classification. *Remote Sensing*, 11(12). https://www.mdpi.com/2072-4292/11/12/1499.

Guo, L., Chehata, N., Mallet, C., Boukir, S., 2011. Relevance of airborne lidar and multispectral image data for urban scene classification using Random Forests. *ISPRS Journal of Photogrammetry and Remote Sensing*, 66(1), 56-66.

Haala, N., Kölle, M., Cramer, M., Laupheimer, D., Mandlburger, G., Glira, P., 2020. Hybrid Georeferencing, Enhancement and Classification of Ultra-High Resolution UAV LiDAR and Image Point Clouds for Monitoring Applications. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, V-2-2020, 727–734.

Hackel, T., Savinov, N., Ladicky, L., Wegner, J. D., Schindler, K., Pollefeys, M., 2017. SEMANTIC3D.NET: A new large-scale point cloud classification benchmark. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, IV-1-W1, 91–98.

Hanocka, R., Hertz, A., Fish, N., Giryes, R., Fleishman, S., Cohen-Or, D., 2019. MeshCNN: A Network with an Edge. *ACM Transactions on Graphics (TOG)*, 38(4), 90.

He, H., Upcroft, B., 2013. Nonparametric semantic segmentation for 3d street scenes. N. Amato (ed.), *IROS2013: IEEE/RSJ International Conference on Intelligent Robots and Systems: New Horizon*, Tokyo Big Sight, Tokyo, Japan.

Hu, S., Liu, Z., Guo, M., Cai, J., Huang, J., Mu, T., Martin, R. R., 2021a. Subdivision-Based Mesh Convolution Networks. *CoRR*, abs/2106.02285. https://arxiv.org/abs/2106.02285.

Hu, W., Zhao, H., Jiang, L., Jia, J., Wong, T., 2021b. Bidirectional Projection Network for Cross Dimension Scene Understanding. *CoRR*, abs/2103.14326. https://arxiv.org/abs/2103.14326.

Jaritz, M., Gu, J., Su, H., 2019. Multi-view PointNet for 3D Scene Understanding. *CoRR*, abs/1909.13603. http://arxiv.org/abs/1909.13603.

Kalogerakis, E., Averkiou, M., Maji, S., Chaudhuri, S., 2017. 3D Shape Segmentation with Projective Convolutional Networks. *CoRR*, abs/1612.02808. http://arxiv.org/abs/1612.02808.

Kalogerakis, E., Hertzmann, A., Singh, K., 2010. Learning 3d mesh segmentation and labeling. *ACM SIGGRAPH 2010 Papers*, SIGGRAPH '10, ACM, New York, NY, USA, 102:1–102:12.

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, 11.

Knott, M., Groenendijk, R., 2021. Towards Mesh-Based Deep Learning for Semantic Segmentation in Photogrammetry. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*.

Kölle, M., Laupheimer, D., Haala, N., 2019. Klassifikation hochaufgelöster LiDAR- und MVS-punktwolken zu monitoringzwecken. *39. Wissenschaftlich-Technische Jahrestagung der OVG, DGPF und SGPF in Wien*, 28, Deutsche Gesellschaft für Photogrammetrie, Fernerkundung und Geoinformation (DGPF) e.V., 692–701.

Laupheimer, D., Haala, N., 2021. Juggling with Representations: On the Information Transfer Between Imagery, Point Clouds, and Meshes for Multi-Modal Semantics. *ISPRS Journal of Photogrammetry and Remote Sensing*, 176, 55-68.

Laupheimer, D., Shams Eddin, M. H., Haala, N., 2020. On the Association of LiDAR Point Clouds and Textured Meshes for Multi-Modal Semantic Segmentation. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, V-2-2020, 509–516.

Lawin, F. J., Danelljan, M., Tosteberg, P., Bhat, G., Khan, F. S., Felsberg, M., 2017. Deep Projective 3D Semantic Segmentation. *CoRR*. http://arxiv.org/abs/1705.03428.

Mallet, C., Bretar, F., Roux, M., Soergel, U., Heipke, C., 2011. Relevance assessment of full-waveform lidar data for urban area classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, 66(6, Supplement), S71-S84. Advances in LIDAR Data Processing and Applications.

Mandlburger, G., Wenzel, K., Spitzer, A., Haala, N., Glira, P., Pfeifer, N., 2017. Improved Topographic Models via Concurrent Airborne LiDAR and Dense Image Matching. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, IV-2/W4, 259–266.

Niemeyer, J., Rottensteiner, F., Soergel, U., 2014. Contextual Classification of LiDAR Data and Building Object Detection in Urban Areas. *ISPRS Journal of Photogrammetry and Remote Sensing*, 87, 152 - 165.

Qi, C. R., Su, H., Mo, K., Guibas, L. J., 2017a. PointNet: Deep learning on point sets for 3D classification and segmentation. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Qi, C. R., Yi, L., Su, H., Guibas, L. J., 2017b. PointNet++: Deep hierarchical feature learning on point sets in a metric space. I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, R. Garnett (eds), *Advances in Neural Information Processing Systems 30*, Curran Associates, Inc., 5105–5114.

Qiao, Y.-L., Gao, L., Yang, J., Rosin, P. L., Lai, Y.-K., Chen, X., 2019. LaplacianNet: Learning on 3D Meshes with Laplacian Encoding and Pooling. abs/1910.14063.

Ramirez, P. Z., Paternesi, C., Gregorio, D. D., Stefano, L. D., 2019. Shooting Labels: 3D Semantic Labeling by Virtual Reality. *arXiv preprint*, abs/1910.05021.

Rothermel, M., Wenzel, K., Fritsch, D., Haala, N., 2012. SURE: Photogrammetric Surface Reconstruction From Imagery. *Proceedings LC3D Workshop*, 8, Berlin.

Rouhani, M., Lafarge, F., Alliez, P., 2017. Semantic Segmentation of 3D Textured Meshes for Urban Scene Analysis. *ISPRS Journal of Photogrammetry and Remote Sensing*, 123, 124 - 139. https://hal.inria.fr/hal-01469502.

Schindler, K., 2012. An Overview and Comparison of Smooth Labeling Methods for Land-Cover Classification. *IEEE Transactions on Geoscience and Remote Sensing*, 50, 4534-4545.

Shilane, P., Min, P., Kazhdan, M., Funkhouser, T., 2004. The Princeton Shape Benchmark. *Shape modeling applications, 2004. Proceedings*, IEEE, 167–178.

Su, H., Maji, S., Kalogerakis, E., Learned-Miller, E., 2015. Multi-View Convolutional Neural Networks for 3D Shape Recognition. *Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV)*, ICCV '15, IEEE Computer Society, Washington, DC, USA, 945–953.

Thomas, H., Qi, C. R., Deschaud, J.-E., Marcotegui, B., Goulette, F., Guibas, L. J., 2019. Kpconv: Flexible and deformable convolution for point clouds. *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*.

Tutzauer, P., Laupheimer, D., Haala, N., 2019. Semantic Urban Mesh Enhancement Utilizing a Hybrid Model. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, IV-2/W7, 175–182.

Weinmann, M., Jutzi, B., Hinz, S., Mallet, C., 2015. Semantic Point Cloud Interpretation Based on Optimal Neighborhoods, Relevant Features and Efficient Classifiers. *ISPRS Journal of Photogrammetry and Remote Sensing*, 105, 286 - 304.

Weinmann, M., Jutzi, B., Mallet, C., 2013. Feature relevance assessment for the semantic interpretation of 3D point cloud data. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, II-5/W2, 313–318.

Wichmann, A., Agoub, A., Kada, M., 2018. ROOFN3D: Deep Learning Training Data for 3D Building Reconstruction. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-2, 1191–1198.

Zhou, Q., 2018. Pymesh. https://github.com/PyMesh/PyMesh.

Zolanvari, S. I., Ruano, S., Rana, A., Cummins, A., da Silva, R. E., Rahbar, M., Smolic, A., 2019. DublinCity: Annotated LiDAR Point Cloud and its Applications. BMVC, 30th British Machine Vision Conference.