2D TO 3D LABEL PROPAGATION FOR THE SEMANTIC SEGMENTATION OF HERITAGE BUILDING POINT CLOUDS

E. Pellis [1*], A. Murtiyoso [3], A. Masiero [1], G. Tucci [1], M. Betti[1], P. Grussenmeyer [2]

[1] Department of Civil and Environmental Engineering (DICEA), University of Florence, 50139 Florence, Italy - (eugenio.pellis, andrea.masiero, grazia.tucci, michele.betti)@unifi.it
[2] Université de Strasbourg, INSA Strasbourg, CNRS, ICube Laboratory UMR 7357, Photogrammetry and Geomatics Group, 67000 Strasbourg, France – pierre.grussenmeyer@insa-strasbourg.fr
[3] Forest Resources Management Group, Institute of Terrestrial Ecosystems, Department of Environmental Systems Science, ETH Zürich, Switzerland - arnadidhestaratri.murtiyoso@usys.eth.ch

Commission II, WG II/8

KEY WORDS: 3D Point Cloud, Semantic Segmentation, Label Transfer, Heritage Buildings

ABSTRACT:

During the last decade, the use of semantic models of 3D buildings and structures kept growing, fostered in particular by the spread of Building Information Models (BIMs), becoming quite popular in several civil engineering and geomatics applications. Nevertheless, semantic model production usually requires quite a lot of human interaction, which may result in quite long and annoying procedures for human operators. The production of 3D semantic models of buildings often takes advantage of already available 3D reconstructions of the considered objects. Given the ever increasing resolution of 3D reconstructions, obtained thanks to the recently developed laser scanners and photogrammetric software, the availability of tools for supporting the automatic or semi-automatic generation of semantic models represents a key step for easing and speeding up the process of semantic model production. In particular, the correct semantic interpretation of the different parts of a 3D point cloud, can be seen as the basic step for the production of a BIM model. The most frequently used methods for point cloud semantic segmentation can be separated in two categories: those directly segmenting the point clouds and those based on the ancillary semantic segmentation of images representing the object of interest, then transferring back the segmentation results to the point cloud. This work focuses on the latter method, considering more specifically the application of heritage building semantic segmentation. To be more specific, this paper investigates the semantic segmentation performance on a set of four heritage buildings, obtained first applying deep-learning based image semantic segmentation and then propagating back the semantic information to the point cloud by means of a voting strategy. The obtained results are quite encouraging, motivating future investigations on improvements of this strategy, in particular when including more buildings in the considered dataset.

1. INTRODUCTION

In the last years there is an increasing interest in the automatic semantic segmentation of 3D point clouds, due to its fundamental role in scene understanding and comprehension in several applications of computer vision, robotic, remote sensing and many others (Zhang et al., 2019). In the Architecture, Engineering and Construction (AEC) sector, Building Information Modelling (BIM) has become a standard design approach, and the use of 3D point clouds is currently the base for as-built BIM model creation (Macher et al., 2015). Modern LiDAR (Light Detection and Ranging) sensors and stereo cameras allow to collect a huge amount of 3D points in short time (Grussenmeyer et al., 2008). On the one hand, this ensures a very detailed spatial representation of the acquired scene. On the other, this also causes quite long processing times when dealing with the raw data, even longer when manual intervention is needed. To leverage this problem, several approaches have been developed aiming at automatizing most of the processing steps. To such aim, machine and deep learning techniques have been extensively investigated during the last years, in particular when dealing with problems related to scene understanding and to the extraction of semantic information, as in this paper (Heipke and Rottensteiner, 2020). Semantic segmentation techniques can be divided into two main groups: projection-based and point-based methods (Xie et al., 2020). Although point-based techniques provide an opportunity for a better understanding of spatial and geometrical information, and they probably are the most promising in the future, the current usability of these methods is quite limited due to the difficulty in acquiring sufficient 3D point labels to properly train a reliable classifier, and due to the high computational cost and the long training time. Instead, the usability of projection-based methods is currently quite good, mostly thanks to the dramatic improvements of neural networks-based image processing. This approach is based on the segmentation of a 2D intermediate representation of the cloud, and then on the reprojection of the extracted labels on the initial cloud. Multiview or image-based approaches leverage on images as intermediate representation of the cloud (Su et al., 2015). Therefore, they allow to exploit the tried-and-tested results obtained by Convolutional Neural Networks (CNNs) on image processing, achieving remarkable results on the semantic segmentation of the representative images (Minaee et al., 2021). A critical step in the image-based semantic segmentation of 3D point clouds is the development of a reliable procedure to re-project the labelling from the 2D representation to the 3D reconstruction. The procedure becomes more challenging when dealing with complex scenarios like the case of heritage

* corresponding author

buildings, in which complex shapes, elements uniqueness and irregular geometries require careful modelling. This paper takes advantage of a previously developed image-semantic segmentation network (Pellis et al., 2022), and it focuses on the development of a reliable 2D to 3D label transfer procedure with the main aim of decreasing the geometric and spatial information loss and improving the overall accuracy of the image-based semantic segmentation workflow. In this paper the procedure will be tested on the case of heritage buildings scenarios, using an ongoing dataset (Pellis et al., 2021) specifically suited for heritage building semantic segmentation.

2. RELATED WORKS

During the last years, several methods have been proposed to face the problem of label projection from 2D images to 3D space to obtain a consistent 3D point cloud segmentation from labelled images. For example, (Wang et al., 2013) design an approach to propagate the pixel-wise image labels from ImageNet to point clouds. In the first step they used *Exemplar SVMs* to over segment individual images into "superpixels", and then propagate their labels onto the visually similar superpixels in the reference images of point cloud. In the second step they used a graphical model to aggregate superpixel label candidates to jointly infer the point cloud labels. Some works on semantic mapping (McCormac et al., 2016), (Hermans et al., 2014) typically aggregated pixel-wise semantic features onto 3D reconstructed surfaces via Bayesian fusion and used Conditional Random Field (CRF) models to regularize the resulting 3D segmentation. In this work (Wang et al., 2019), the authors present *Label Diffusion Lidar Segmentation* (LDLS), a method for instance segmentation of 3D point clouds which leverages a pretrained 2D image segmentation model. They obtain 2D segmentation prediction by applying Mask-RCNN, and then link the image to a 3D lidar point cloud by building a graph of connections among 3D points and 2D pixels. (Zhang et al., 2018) addressed the issue of the semantic segmentation of large-scale 3D scenes by fusing 2D images and 3D point clouds. According to this work the preliminary segmentation results with 2D images obtained by a DeepLab-Vgg16 based model, are mapped to 3D point clouds according to the coordinate relationship between the images and the point cloud calculated with DLT algorithm. More recently, (Genova et al., 2021) proposed a novel network 2D3DNet, that uses multi-view fusion to make best-guess semantic labels for as many 3D points as possible via back-projection and voting from labels of the corresponding pixels. (Mascaro et al., 2021) presented *Diffuser*, a novel framework that leverages 2D semantic segmentation to produce a consistent 3D segmentation. They formulate the 3D segmentation task as transductive label diffusion problem on a graph, where multi-view and 3D geometric proprieties are used to propagate semantic labels from the 2D space to the 3D map. They show a significant accuracy compared to probabilistic fusion methods. The approach developed in (Lertniphonphan et al., 2018), propagate object label from 2D image to a sparse point cloud by matching a group of points that corresponds to the area within the 2D bounding box in the image. The method was used for producing training data, and it demonstrates that the label propagation can be used to train a classifier with a good average precision. In the specific context of building segmentation (Murtiyoso et al., 2021) proposed an approach for the segmentation of 3D building façade based on orthophoto. The XY coordinates of each pixel in the orthophoto was used to determine the corresponding planimetric coordinates of the point in the point cloud and finally a winner-takes-all approach was applied to annotate the 3D points with the respective 2D pixel class. In a more recent work (Murtiyoso et al., 2022) introduced semantic classification at the beginning of the classical photogrammetric workflow in order to automatically create a classified dense point cloud. In this regard, several image masks obtained by a trained neural network are employed during dense image matching in order to constraint the process into the respective classes. In the same context (Stathopoulou and Remondino, 2019) proposed a semantic photogrammetry workflow, in which the label back-projection is based on the projection matrix P which connects the 3D with the 2D space. The segmented images are automatically generated using neural networks, and then the labels are used as constraints in the photogrammetric process. Giving the correspondence, all the images contribute to the labelling projection on the cloud with a weighted winner procedure.

3. DEVELOPED METHODOLOGY

The proposed methodology aims at projecting the labels, predicted by a deep learning-based image semantic classifier on a set of N 2D images, on a 3D point cloud. The interior and exterior parameters of the images imputed in the deep-learning classifier are assumed to be known: despite such parameters could be computed aside of the point cloud generation, their availability comes for free when the point cloud is the outcome of a photogrammetric reconstruction procedure, and the images inputted in the classifier are taken among those used in the reconstruction. Hence, this could be considered as a quite ideal working condition for the proposed method.

In accordance with the above consideration, hereafter the considered images are assumed to have already been aligned, and the exterior parameters are assumed to be expressed in a reference system compatible with the point cloud one.

The procedure starts with the image segmentation step: despite any proper image semantic segmentation procedure could be viable in this step, the deep-learning method proposed in (Pellis et al., 2022) has been used in this work, providing N semantically segmented images $\{I_j\}_{j=1,\dots,N}$ as output.

Then, the labels of the N predicted images are properly transferred to the point cloud, as described in the following.

1. 3D points of the cloud are projected on the N images, by means of the known interior and exterior camera parameters.
2. For each image I_j, each point class is assessed, if visible.
3. For each point, the mostly voted class is selected.

Let (u_j,v_j) be the pixel coordinates of the projection of point p on the image I_j. A straightforward implementation of step 2 is the assignation of the label of pixel (u_j,v_j) in I_j (if inside the image extent) as its vote to point p class.

Despite being very simple, such a strategy does not take into account of the obstructions, leading to unreliable outcomes in complex scenarios: the implementation of an effective procedure to check obstructions is of vital importance for ensuring a good performance of the overall algorithm in a wide range of working conditions.

Assume that the point cloud density is sufficiently high to ensure that at least one 3D point is projected in all the adjacent pixels, in image I_j, describing the same object surface. Down-sampling the image size, or, equivalently, enlarging the pixel size, could be necessary in order to ensure the validity of such assumption.

According to the above hypothesis, at least two points should be projected on the same pixel (u_j,v_j) when an obstruction occurs. When such event is detected, a simple check on the distance between the camera and the points projected on the same pixel is used in order to determine if any of such points probably obstructs the others. Image I_j votes only for the non-obstructed points. The main advantages of such procedure are the

implementation simplicity and the quite effectiveness in most of the examined conditions. Nevertheless, a more complex strategy will be considered in our future investigations in order to improve the semantic segmentation results in quite critical conditions.

4. DATASET

To test the proposed procedure, the dataset presented in (Pellis et al., 2021), currently composed by three heritage buildings, has been used. For each building several data types are available: (i) the Terrestrial Laser Scanner (TLS) cloud with the corresponding ground-truth segmentation, (ii) the photogrammetric cloud with the corresponding ground-truth segmentation, and (iii) the RGB images of the photogrammetric survey with the corresponding pixel-wise ground-truth segmentation. Since the images were previously used for the creation of the photogrammetric cloud, the internal and the external camera parameters are also available. Figure 1 shows some images of the buildings in the dataset.

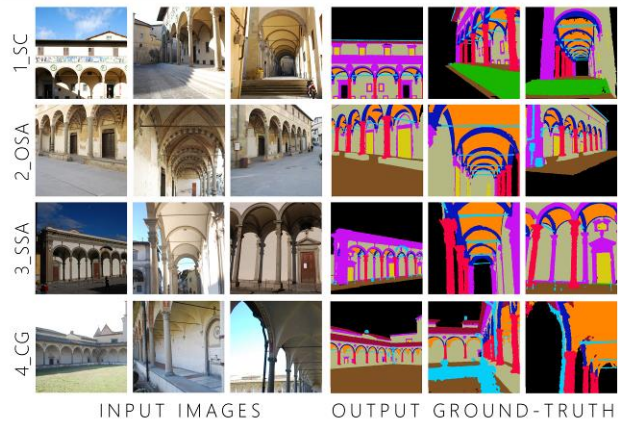

Figure 1. Examples of the images in the dataset with the corresponding ground-truth.

The segmentation classes considered in this dataset are structured following the guidelines and the standards of ARCHdataset (Matrone et al., 2020) an existing benchmark for point cloud semantic segmentation. The classes refer to the IFC file format, to CityGML (LOD3/4) and to ATT (Art and Architecture Thesaurus). They include 11 categories: *0_arch, 1_column, 2_moulding, 3_floor, 4_window/door, 5_wall, 6_stair, 7_vault, 8_roof, 9_other and 10_background.* The distribution of the points among such classes in i) the LiDAR point clouds, ii) the photogrammetric point clouds and iii) in the images of the dataset is shown in Figure 2.

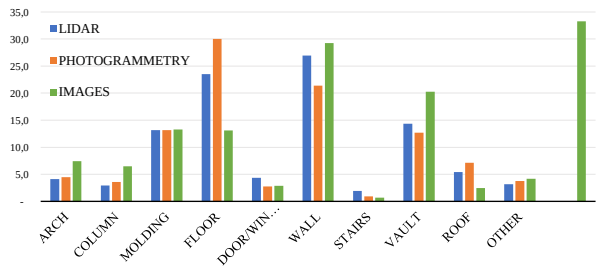

Figure 2. Class percentage distribution for the TLS clouds (blue), for the photogrammetric clouds (orange), and for the images (green).

5. RESULTS

Aiming at checking the performance of the proposed approach in several working conditions, some tests for each of the three available buildings of the dataset have been run. First, a deep neural network has been trained for each building, splitting the images of the same building in training set, validation set and test set. Secondly, we used the predicted pixel-wise labelling of the test set to project the features from the images to the corresponding photogrammetric cloud. To assess the performance of the reprojection procedure, we compared the obtained labelled cloud with the ground-truth point cloud, and we evaluated the performance degradation comparing the results with the accuracy obtained by the neural network on the 2D segmentation.

To evaluate the performance of our models we used two evaluation metrics: the Global Accuracy (GA), and the mean Intersection Over Union (mIoU) defined in the equations below:

$$GA = \frac{\sum_i n_{ii}}{\sum_i t_i} \quad (1)$$

$$mIoU = \frac{1}{n_c l} \sum_i \frac{n_{ii}}{(t_i + \sum_j n_{ji} - n_{ji})} \quad (2)$$

where $n_c l$ = number of classes included in ground truth
n_{ij} = number of pixels of class i predicted to belong class j
t_i = total number of pixels of class i in ground truth

For each model, the confusion matrix will be shown as well, in order to provide a more in-depth analysis of the semantic segmentation performance.

In the next sections we are going to show at first the results for the image segmentation (5.1), and secondly the results for the projection procedure (5.2).

5.1 Image Semantic Segmentation

Image semantic segmentation is a key step in many computer vision applications, and hence several approaches to implement it have been proposed in literature. Over the past few years, however, CNNs have yielded a new generation of models with a remarkable performance improvement. In our tests we exploited one of the most prominent CNN-based models, DeepLabv3+ (Chen et al., 2018). We used a pretrained version of the network on the ImageNet database (Deng et al., 2009) with ResNet-18 (He et al., 2015) as base classification architecture. The testing dataset (Pellis et al., 2021) is still in progress, and it still lacks of a sufficient variability in the images and building typologies to well-generalize a complete unseen scenario.

Nevertheless, some tests, varying the complexity of the goal, have been performed to check the label prediction ability of the network, as shown below.

For each building we randomly shuffled all the image, and we randomly split them in training set (60%), validation set (20%) and test set (20%) (Dobbin and Simon, 2011) Then, the labels predicted in the test set were used in the back-projection procedure. Table 1 shows the distribution of the images in training, validation and test sets for each of the considered sets.

Building	N° of Image	TrainingSet	ValidationSet	TestSet
1_SC	748	448	150	**150**
2_OSA	755	453	151	**151**
3_SSA	473	283	95	**95**

Table 1. Number of images in Training Set, Validation Set and Test Set for each building.

The various models were trained for 30 epochs, using Stochastic Gradient Descend with Momentum (SGDM) as optimizer, and with the same hyperparameters for each case study. The image segmentation performance obtained on the test sets of each of the buildings is reported in Table 2.

Building	GlobalAccuracy	mIoU	mBFScore
1_SC	0.90	0.76	0.77
2_OSA	0.89	0.78	0.78
3_SSA	0.87	0.62	0.72

Table 2. Image semantic segmentation results on the Test Set of the three buildings in the dataset.

The obtained results are quite satisfactory, in particular for the first two buildings, yielding a GA around 90% and a mIoU around 80%. The reader is referred to (Pellis et al., 2022) for more details on the training phase and on the obtained image segmentation results.
Figure 3 reports some comparisons between the ground truth and the predicted image segmentation on the test sets.

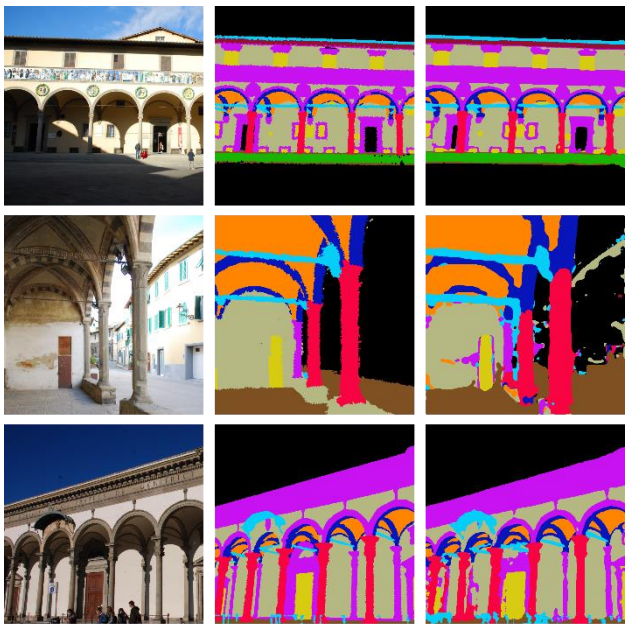

Figure 3. DeepLablv3+ image semantic segmentation results on the test sets: a) input RGB images, b) ground truth, c) prediction.

5.2 Labelling Projection

In this section we are going to show more in detail the results of the labelling procedure for each of the three buildings of the dataset. For each building, all the labelled images of the test set outputted by the neural network have been used as input in the label back-projection procedure.
A cleaned, denoised and subsampled version of the photogrammetric point cloud has been inputted in the back-projection procedure, along with the predicted image labels. Among the cleaning operations, it is worth to notice that the "background" points were removed from the cloud.
Examples of the graphical outcomes of the back-projection procedures are shown in Figure 4, 6 and 8, for the different buildings. The obtained numerical results, in terms of GA and mIoU, are reported in Table 3, 4 and 5, whereas the corresponding confusion matrices are shown in Figure 5, 7 and 9.
It is worth to notice that, as a consequence of the implemented way to deal with occlusions, a portion of the points is not classified by the proposed algorithm. Since most of the points are classified, the unlabelled points could be reconsidered for classification as a further step of the proposed back-projection label transferring procedure, as will be investigated in our future works.
In accordance with the above consideration, the results limited to only the classified points (second row in Table 3, 4 and 5) are those considered more relevant here.

1_SC Spedale del Ceppo

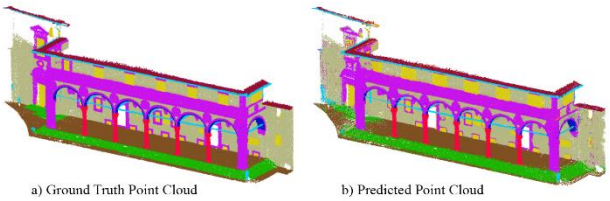

Figure 4. Comparison between a) the Ground Truth point cloud, b) the Predicted point cloud.

Reference Points	GlobalAccuracy	mIoU	% Labelled Points
All points	0.76	0.61	100
Only Classified	**0.87**	**0.70**	87

Table 3. Back-projection results for 1_SC.

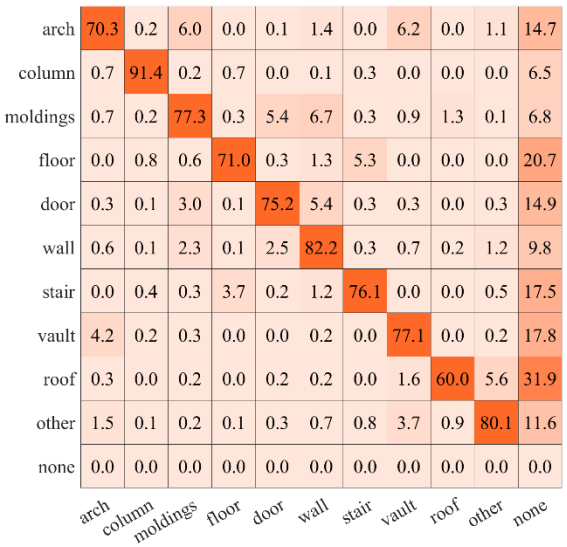

Figure 5. Confusion Matrix for 1_SC back-projection.

2_OSA Ospedale di Sant'Antonio

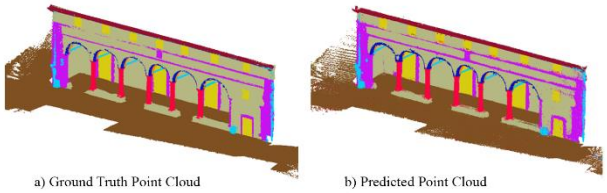

Figure 6. Comparison between a) the Ground Truth point cloud, b) the Predicted point cloud

Reference Points	GlobalAccuracy	mIoU	% Labelled Points
All points	0.60	0.52	100
Only Classified	**0.91**	**0.72**	67

Table 4. Back-projection results for 2_OSA.

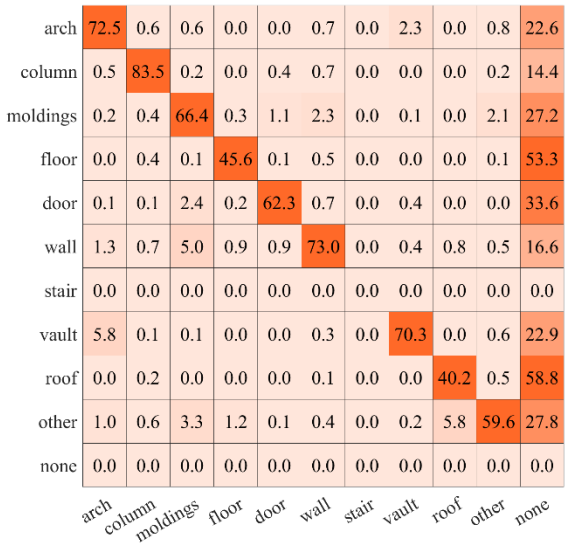

Figure 7. Confusion Matrix for 2_OSA back-projection.

3_SSA Basilica della Santissima Annunziata

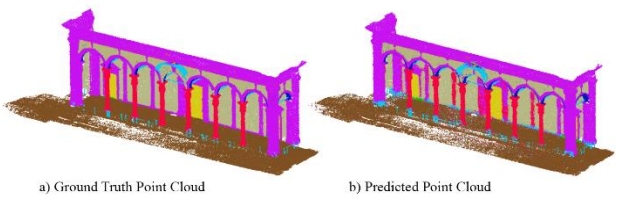

Figure 8. Comparison between a) the Ground Truth point cloud, b) the Predicted point cloud.

Reference Points	GlobalAccuracy	mIoU	% Labelled Points
All points	0.72	0.52	100
Only Classified	**0.88**	**0.61**	82

Table 5. Back-projection results for 3_SSA.

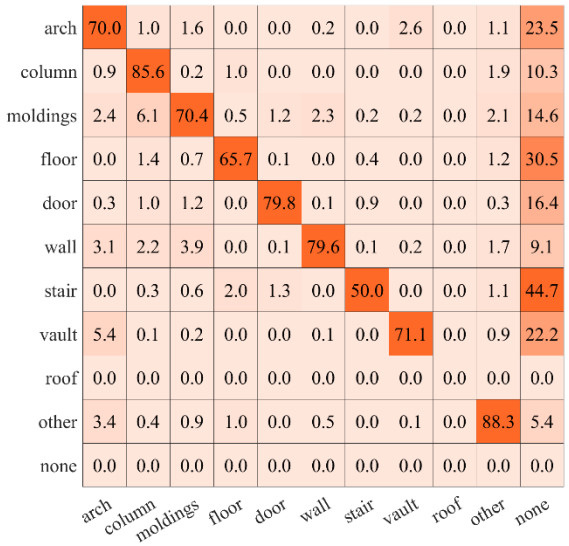

Figure 9. Confusion Matrix for 3_SSA back-projection.

6. DISCUSSION

The proposed label propagation method obtained quite remarkable results on the three case studies, in particular when considering only the labelled points, when compared with the image segmentation outcomes.
To be more precise, Figure 10 shows a comparison between the image and the point cloud semantic segmentation performance. The resulting GA and mIoU are quite similar in all the considered cases: the obtained results reveal that the quality of the obtained point cloud semantic segmentation is mostly related to the that of the image segmentation. As a consequence of such observation, an increase in the image segmentation accuracy should directly correspond to an improvement on the point cloud segmentation performance.

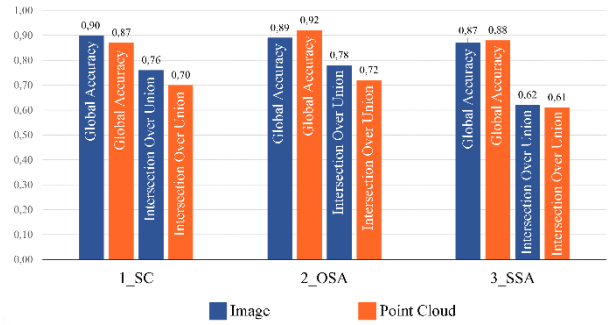

Figure 10. Comparison between GA e mIoU on image segmentation (blue) and on the point cloud segmentation (orange).

A close look to the confusion matrices (discarding unlabelled points) shows the absence of remarkably bad-segmented classes, although the performance may change from case to case. This is also confirmed by Figure 11, where the sum of GA (blue bars) and unlabelled point percentage (black bars), derived by considering all the three buildings, is quite close to 100% for almost all the classes

GA values are quite balanced in the classes, with lower values for "roof" and "stair", which are also certain of those less represented in the dataset (see Figure 2).
The percentage of unlabelled points is quite widely variable, with large values on the more frequently obstructed classes. The introduction of an additional step in the label propagation procedure, in order to ensure the classification of most the currently unlabelled points, will be considered in our future works, as previously mentioned.

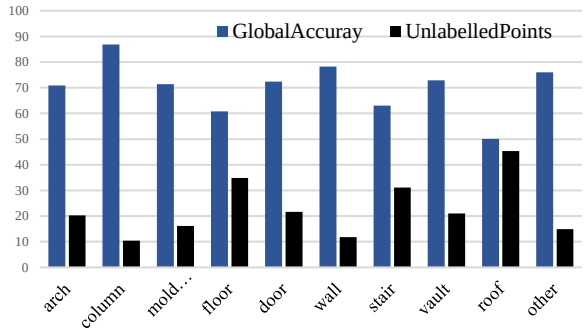

Figure 11. Global Accuracy (blue) and unlabelled points percentage (black) for each class.

Finally, for comparison with the results presented here, Table 6 shows those obtained with the masking-based methodology described in (Murtiyoso et al., 2022) and carried out in (Pellis et al., 2022).

Building	GlobalAccuracy	mIoU
1_SC	0.67	0.52
2_OSA	0.72	0.45
3_SSA	0.75	0.44

Table 6. Masking-based methodology results.

Despite the masking method works well for background removal and for building façade classification, the label propagation method considered here outperforms the masking-based one on our dataset, highlighting the still challenging use of semantically enriched reconstruction methods in complex scenarios.

7. CONCLUSION

In this paper we presented a procedure for the label propagation of semantic classifications from 2D images to 3D point clouds, tested on a dataset composed by three heritage building. The first results have shown a quite remarkable performance of the proposed back-projection approach, ensuring a classification accuracy similar to the image segmentation performance. Overall, the proposed procedure outperformed masking-based methods on the label propagation, but not influencing the 3D point cloud generation procedure, hence, for instance, not ensuring any cleaning effect that could come as a result of the masking methods. Future investigations will be dedicated in particular to reduce the percentage of unlabelled points. For what concerns the point cloud semantic segmentation results, the obtained results revealed that the bottleneck of the entire workflow is the neural network-based image classification performance, which is negatively influenced by the intrinsic complexity of buildings in the heritage scenario. Increasing the variability inside of the dataset, both in terms of number of buildings and of images, is expected to have a positive impact on this aspect. Data augmentation and synthetic data generation will also be considered in our future works in order to increase the generalization capability of the image semantic segmentation network.

REFERENCES

Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H., 2018. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation, in: IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR).

Deng, J., Dong, W., Socher, R., Li, L.-J., Kai Li, Li Fei-Fei, 2009. ImageNet: A large-scale hierarchical image database, in: IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR). Institute of Electrical and Electronics Engineers (IEEE), pp. 248–255. https://doi.org/10.1109/cvpr.2009.5206848

Dobbin, K.K., Simon, R.M., 2011. Optimally splitting cases for training and testing high dimensional classifiers. BMC Medical Genomics 4. https://doi.org/10.1186/1755-8794-4-31

Genova, K., Yin, X., Kundu, A., Pantofaru, C., Cole, F., Sud, A., Brewington, B., Shucker, B., Funkhouser, T., 2021. Learning 3D Semantic Segmentation with only 2D Image Supervision, in: 2021 International ViConference on 3D Vision (3DV). https://doi.org/10.1109/3dv53792.2021.00046

Grussenmeyer, P., Landes, T., Voegtle, T., Ringle, K., 2008. Comparison Methods Of Terrestrial Laser Scanning, Photogrammetry And Tacheometry Data For Recording Of Cultural Heritage Buildings, in: International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives.

He, K., Zhang, X., Ren, S., Sun, J., 2015. Deep Residual Learning for Image Recognition, in: Proceedings of the IEEE Conference On Computer Vision and Pattern Recognition. pp. 770–778.

Heipke, C., Rottensteiner, F., 2020. Deep learning for geometric and semantic tasks in photogrammetry and remote sensing. Geo-Spatial Information Science 23, 10–19. https://doi.org/10.1080/10095020.2020.1718003

Hermans, A., Floros, G., Leibe, B., 2014. Dense 3D Semantic Mapping of Indoor Scenes from RGB-D Images, in: IEEE International Conference on Robotics and Automation (ICRA).

Lertniphonphan, K., Satoshi, K., Kazuyuki, T., Hiromasa, Y., 2018. 2D to 3D Label Propagation for Object Detection in Point Cloud, in: IEEE International Conference on Multimedia & Expo Workshops (ICMEW). pp. 1–6.

Macher, H., Landes, T., Grussenmeyer, P., 2015. Point clouds segmentation as base for as-built BIM creation, in: ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences. Copernicus GmbH, pp. 191–197. https://doi.org/10.5194/isprsannals-II-5-W3-191-2015

Mascaro, R., Teixeira, P., Teixeira, L., Chli, M., 2021. Diffuser: Multi-View 2D-to-3D Label Diffusion for Semantic Scene Segmentation, in: IEEE International Conference OnRobotics and Automation (ICRA). https://doi.org/10.3929/ethz-b-000484229

Matrone, F., Lingua, A., Pierdicca, R., Malinverni, E.S., Paolanti, M., Grilli, E., Remondino, F., Murtiyoso, A., Landes, T., 2020. A Benchmark for Large-Scale Heritage Point Cloud Semanti Segmentation, in: International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives. International Society for Photogrammetry and Remote Sensing, pp. 1419–1426. https://doi.org/10.5194/isprs-archives-XLIII-B2-2020-1419-2020

McCormac, J., Handa, A., Davison, A., Leutenegger, S., 2016. SemanticFusion: Dense 3D Semantic Mapping with Convolutional Neural Networks, in: IEEE International Conference on Robotics and Automation (ICRA).

Minaee, S., Boykov, Y.Y., Porikli, F., Plaza, A.J., Kehtarnavaz, N., Terzopoulos, D., 2021. Image Segmentation Using Deep Learning: A Survey. IEEE Transactions on Pattern Analysis and Machine Intelligence. https://doi.org/10.1109/TPAMI.2021.3059968

Murtiyoso, A., Lhenry, C., Landes, T., Grussenmeyer, P., Alby, E., 2021. Semantic segmentation for building façade 3D point cloud from 2D orthophoto images using transfer learning, in: International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives. International Society for Photogrammetry and Remote Sensing, pp. 201–206. https://doi.org/10.5194/isprs-archives-XLIII-B2-2021-201-2021

Murtiyoso, A., Pellis, E., Grussenmeyer, P., Landes, T., Masiero, A., 2022. Towards Semantic Photogrammetry: Generating Semantically Rich Point Clouds from Architectural Close-Range Photogrammetry. Sensors 22. https://doi.org/10.3390/s22030966

Pellis, E., Masiero, A., Tucci, G., Betti, M., Grussenmeyer, P., 2021. Assembling an Image and Point Cloud Dataset for Heritage Buildings Semantic Segmentation. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XLVI-M-1–2021, 539–546. https://doi.org/10.5194/isprs-archives-xlvi-m-1-2021-539-2021

Pellis, E., Murtiyoso, A., Masiero, A., Tucci, G., Betti, M., Grussenmeyer, P., 2022. An Image-Based Deep Learning workflow for 3D Heritage Point Cloud Semantic Segmentation. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XLVI-2/W1-2022, 429–434. https://doi.org/10.5194/isprs-archives-XLVI-2-W1-2022-429-2022

Stathopoulou, E.K., Remondino, F., 2019. Semantic Photogrammetry - Boosting Image-based 3D Reconstruction with Semantic Labeling. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences 42, 685–690. https://doi.org/10.5194/isprs-archives-XLII-2-W9-685-2019

Su, H., Maji, S., Kalogerakis, E., Learned-Miller, E., 2015. Multi-view convolutional neural networks for 3D shape recognition, in: Proceedings of the IEEE International Conference on Computer Vision. pp. 945–953. https://doi.org/10.1109/ICCV.2015.114

Wang, B.H., Chao, W.L., Wang, Y., Hariharan, B., Weinberger, K.Q., Campbell, M., 2019. LDLS: 3-D Object Segmentation Through Label Diffusion from 2-D Images. IEEE Robotics and Automation Letters 4, 2902–2909. https://doi.org/10.1109/LRA.2019.2922582

Wang, Y., Ji, R., Chang, S.F., 2013. Label propagation from image net to 3D point clouds. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition 3135–3142. https://doi.org/10.1109/CVPR.2013.403

Xie, Y., Tian, J., Zhu, X.X., 2020. Linking Points with Labels in 3D: A Review of Point Cloud Semantic Segmentation. IEEE Geoscience and Remote Sensing Magazine 8, 38–59. https://doi.org/10.1109/MGRS.2019.2937630

Zhang, J., Zhao, X., Chen, Z., Lu, Z., 2019. A Review of Deep Learning-Based Semantic Segmentation for Point Cloud. IEEE Access 7, 179118–179133. https://doi.org/10.1109/ACCESS.2019.2958671

Zhang, R., Li, G., Li, M., Wang, L., 2018. Fusion of images and point clouds for the semantic segmentation of large-scale 3D scenes based on deep learning. ISPRS Journal of Photogrammetry and Remote Sensing 143, 85–96. https://doi.org/10.1016/j.isprsjprs.2018.04.022