Semantic segmentation for building façade 3D point cloud from 2D orthophoto images using transfer learning

Arnadi Murtiyoso, Camille Lhenry, Tania Landes, Pierre Grussenmeyer, Emmanuel Alby

SEMANTIC SEGMENTATION FOR BUILDING FAÇADE 3D POINT CLOUD FROM 2D ORTHOPHOTO IMAGES USING TRANSFER LEARNING

A. Murtiyoso*, C. Lhenry, T. Landes, P. Grussenmeyer, E. Alby

Université de Strasbourg, INSA Strasbourg, CNRS, ICube Laboratory UMR 7357, Photogrammetry and Geomatics Group, 67000 Strasbourg, France - (arnadi.murtiyoso, camille.lhenry, tania.landes, pierre.grussenmeyer, emmanuel.alby)@insa-strasbourg.fr

Commission II, WG II/3

KEY WORDS: point cloud, semantic segmentation, orthophoto, photogrammetry, façade, transfer learning, deep learning

ABSTRACT:

The task of semantic segmentation is an important one in the context of 3D building modelling. Indeed, developments in 3D generation techniques have rendered the point cloud ubiquitous. However pure data acquisition only captures geometric information and semantic classification remains to be performed, often manually, in order to give a tangible sense to the 3D data. Recently progress in computing power also opened the way for massive application of deep learning methods, including for semantic segmentation purposes. Although well established in the processing of 2D images, deep learning solutions remain an open question for 3D data. In this study, we aim to benefit from the vastly more developed 2D semantic segmentation by performing transfer learning on a photogrammetric orthoimage. The neural network was trained using labelled and rectified images of building façades. Another programme was then written to permit the passage between 2D orthoimage and 3D point cloud. Results show that the approach worked well and presents an alternative to help the automation process for point cloud semantic segmentation, at least in the case of photogrammetric data.

1. INTRODUCTION

The segmentation of 3D point cloud into different classes is an important task to support various workflows, for example the generation of 3D GIS and BIM (Bassier et al., 2017). This has been done using various approaches which can generally be divided into algorithmic or machine learning-based methods (Murtiyoso et al., 2020). Algorithmic approaches have shown to generate good results (Maalek et al., 2019) while machine-learning (ML) based classification methods have also showed promising results (Malinverni et al., 2019). ML techniques are more robust against noise and occlusions, but require training data which may not always be available.

In the field of machine learning, deep learning (DL) as a subset has gained much attention in recent years aided by the recent developments in computing power and availability of large training datasets. DL employs multiple hidden layers of self-optimising artificial neurons, as opposed to shallow neural networks. It is most prominently used in 2D image classification, but recent developments have progressed towards semantic, instance, and even panoptic segmentation (Kirillov et al., 2019) in 3D data classification. A DL framework for 3D point cloud saw a breakthrough with the PointNet++ (Qi et al., 2017b) and subsequently other frameworks. That being said, 2D semantic segmentation is nowadays a stable method that may be deployed easily.

In this paper, we develop an approach to semantically segment building façades using DL by deploying pre-existing and pre-trained networks and tuning them to our requirements; a process known as transfer learning. In order to benefit fully from pre-existing networks, the approach performs semantic segmentation on 2D orthophoto images generated by photogrammetry. A previous research attempted to use fuzzy logic in performing image classification (Neusch and Grussenmeyer, 2003). With a similar purpose, Grilli et al. (2018) performed supervised segmentation on orthophotos and UV textures of photogrammetric 3D models, but using an ML-based approach. In this paper, we present a DL-based transfer learning approach for performing a similar task.

The developed approach deploys a pre-trained ResNet-18 network (He et al., 2016) to create a DeepLabv3+ network (Chen et al., 2018) by training it on a dataset of several hundreds of labelled rectified images of building façades. The trained network was then applied to semantically segment an orthoimage (thus 2.5D) of a case-study building. From this image, a function was employed to back-project the pixel coordinates into the 3D point cloud, thus extracting the point cloud of each class.

2. RELATED WORK

Semantic analysis firstly of 2D and more recently of 3D scenes has become an increasingly studied topic in various applications, such as photogrammetry, remote sensing, computer vision and robotics (Heipke and Rottensteiner, 2020). Input data might be images, point clouds or textured meshes. Several approaches have been developed to automate this task with promising results, such as algorithmic (Maalek et al., 2019) and machine learning (Matrone et al., 2020) approaches. At the same time, deep learning and neural networks are becoming more popular due to increasing computational capacity and growing number of available databases.

Neural networks are generally used for several tasks: classification, semantic segmentation, instance segmentation

* Corresponding author

and more recently, panoptic segmentation. Classification refers to the global prediction of an input: it can be the class of an object on an image, or the nature of a whole point cloud. The semantic segmentation provides a fine understanding of a scene by assigning a label to each pixel of an image or to each point of a point cloud. To go even further in the comprehension of the environment, instance segmentation distinguishes the instances of objects belonging to the same semantic class. Thus, it can be interpreted as a combination of object detection and semantic segmentation (Hafiz and Bhat, 2020). A final task which achieves a rich and complete knowledge of a scene is called panoptic segmentation. It combines semantic and instance segmentation, providing a per-pixel or per-point label that combines class and instance information (Kirillov et al., 2019).

The rapid development of 3D acquisition techniques (lidar, mobile mapping, etc.) has enabled the production of more and more data at lower cost and simultaneously in large quantities. As a consequence, researchers have begun to focus on the application of neural networks for point cloud processing, leading to the emergence of new architectures. These algorithms can be divided into three types: projection-based, discretisation-based and point-based methods (Guo et al., 2020).

The first two techniques use an intermediate representation of the point cloud. Originally unstructured in nature, the common idea is to order it in a manner that allows the use of convolution operators. Discretisation-based methods consist in transforming the point cloud into a regular voxel grid. Projection-based methods project the point cloud in several 2D views to extract feature maps and perform predictions on it. The result is then projected onto the point cloud, using a depth map (Guo et al., 2020). However, these methods are limited: switching to an intermediate representation involves a loss of information. Moreover, the result depends on the choice of projections or voxel sampling of the point cloud. Finally, these techniques are computationally very expensive (Qi et al., 2017a).

Confronted with these challenges, approaches dealing directly with point clouds have been developed. The reference architectures are PointNet (Qi et al., 2017a) and its improved version PointNet++ (Qi et al., 2017b) which considers local features. According to the authors, it is more efficient, fast and robust than methods based on intermediate representations.

The quality of the database needed to train the neural network is a crucial issue, as it will directly influence its performance. One of the main challenges of using neural networks for 3D tasks is the lack of reference databases, compared to those available for image processing. Moreover, manual labelling of a point cloud database is a very time-consuming process (Zolanvari et al., 2019). Faced with this problem, we argue that it is interesting to look at projection-based solutions, where labelled databases are abundantly available. This argument is also backed by other research results which show that object classes concerning building façade such as windows and doors are more difficult to detect using a purely 3D approach (Malinverni et al., 2019, Pierdicca et al., 2020). This problem mainly rose from the lack of training data for building opening classes, as well as the difficulty in distinguishing different classes in generally homogeneous and with reduced relief surfaces such as façades.

Automatic image segmentation is a research topic that has been widely covered in the artificial intelligence community, especially with the use of convolutional neural networks (Kaushik and Kumar, 2019). Taking advantage of the rapid increase in the number of labelled image databases and the improvement of the computing capacities of graphics processor units, this type of architecture has given state-of-the-art results in various fields such as medical imaging, road transportation, product quality monitoring, etc. (Meyer et al., 2018).

There exists many architectures for CNNs; however, they are all based on three similar layers: convolutional, pooling and fully connected layer. The convolutional layer aims to learn the characteristics of the images, called feature maps. The pooling layer allows to reduce the dimensionality of the feature maps by extracting the most relevant features. After a certain number of convolutional and pooling layers, a fully connected layer is implemented to perform the segmentation task (Gu et al., 2017).

The architecture chosen for this study is the DeepLabv3+ neural network. It was used in Tang et al. (2020) where it was associated with a Faster-R-CNN architecture for liver segmentation on medical images. In the cited paper, DeepLabv3+ achieved better performance than other state-of-art methods used in this domain.

The aim of this paper is therefore to utilise the DeepLabv3+ network and train it on a database of labelled building façade images. The trained network will thereafter be deployed on 2D orthoimages generated by photogrammetry, before back-projected to the 3D space later on.

3. METHODOLOGY

The deployment of the developed approach requires three inputs: (1) a georeferenced orthophoto image of the object generated by photogrammetry, (2) a depth map with 2.5D depth in the same coordinate system as the orthophoto, and (3) a point cloud of the object. The point cloud may come from any source; however it should be georeferenced in the same system as the orthophoto and the depth map. In this paper, these three inputs were acquired from terrestrial photogrammetry, but other methods of orthophoto, depth map, and point cloud generation may be envisaged e.g. mobile laser scanner (MLS) or static terrestrial laser scanners (TLS). As can be consulted in Figure 1, in general the approach consists of three steps: generation of input using photogrammetry, semantic segmentation using DL, and back-projection from 2D to 3D.

3.1 Input data generation and pre-processing

The first step of the study involves the creation of the three required input data. For the purposes of these experiments, the main façade of the Zoological Museum in Strasbourg, France was used as the test data. The façade was reconstructed using terrestrial photogrammetry using a Canon EOS 6D camera. The 3D point cloud was generated through dense matching using the software Agisoft Metashape.

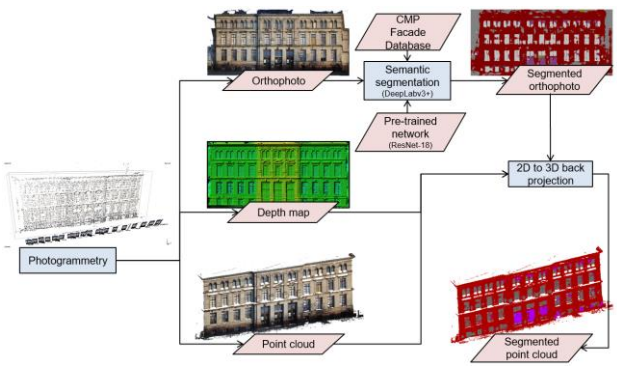

Figure 1. General workflow of the developed approach.

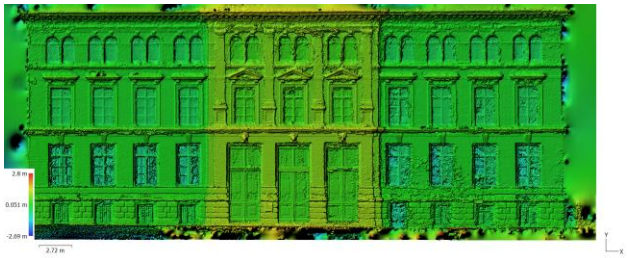

Figure 2. Depth map of the Zoological Museum façade generated by terrestrial photogrammetry used as input.

Figure 3. Corresponding orthophoto to the depth map in Figure 2 created also from photogrammetry.

As the images were taken from an average camera-to-object distance of 10 m, this yields a Ground Sampling Distance (GSD) of roughly 4 mm. The point cloud used in this regard was generated using the "High" quality in Metashape, which implies the subsampling of the raw images with a factor of 2 for the dense matching process; the theoretical spatial resolution of the point cloud is therefore 8 mm.

The point cloud was scaled using coordinates from control points located on the façade and measured for the purposes of the research project. Furthermore, in order to facilitate the algorithm down the pipeline, the 3D photogrammetric network was rotated on the Z axis in such way that the façade is perpendicular to the Y axis. The transformation was computed automatically by identifying the point cloud's major axes using the Principal Component Analysis (PCA) method.

Using the rotated point cloud, a depth map (Figure 2) was computed from an arbitrary reference plane. Consequently, an orthophoto (Figure 3) was computed using this depth map with a set pixel resolution of 1 cm. The generated point cloud, depth map, and orthophoto are all in the same rotated coordinate system as required by the semantic segmentation algorithm.

3.2 Deep learning training and deployment

As has been previously established, the DL part of the method is based on the transfer learning approach. In this study, a DeepLabv3+ network was created from a pre-trained ResNet-18 network (He et al., 2016). The choice of ResNet-18 as the initial network was empirical; more experiments and tests need to be performed to assess the potential results from other pre-trained networks.

The pre-trained network was thereafter augmented using another dataset of labelled rectified images of façades prepared by the Center for Machine Perception (CMP) of the Czech Technical University, Prague (Tyleček and Šára, 2013). This dataset consists of 606 rectified images of building façades (Figure 4) of various architectural styles, although the majority seems to belong to modern and/or European style buildings.

Figure 4. Sample rectified images of building façades from the CMP database. The complete database comprises of 606 labelled images from around the world.

The images were classified into 12 classes, of which only six ("shops", "pillar", "door", "window", "façade", and "background") are retained in this study with other classes merged into these six. The CMP façade dataset can be accessed here: https://cmp.felk.cvut.cz/~tylecr1/facade/ (last accessed 2 February 2021). The choice of the classes included in the training was based on their existence in the test image of the Zoological Museum's façade orthophoto. While no free-standing pillars exist in the test dataset, several engaged columns can be observed on the second floor. "Shops" in this regard refer to business signs and plaques.

For the purpose of the training, the labelled image dataset was randomly divided into training (80%) and validation data (20%). The training was performed for 100 epochs yielding a validation accuracy of 79.65%. Once trained, the network was deployed to predict the classes of each pixel in the input orthophoto image. In these experiments, a spatial pixel resolution of 1 cm for the orthophoto and depth map was used.

3.3 Back-projection from 2D to 3D

The last step of the developed method involves the back-projection of the classified pixels into the 3D point cloud. Since all inputs are in the same coordinate system, this process was quite straightforward. The XY coordinates of each pixel in the orthophoto was used to determine the corresponding planimetric coordinates (with regards to the façade plane) of the points in the point cloud.

The identical resolution of the orthophoto and depth map means that the depth value from the corresponding pixel in the depth map can then be directly correlated and therefore extracted to obtain the depth element for each orthophoto pixel.

Using this method, each pixel divides the point cloud into voxel-like cuboids with the same XY dimensions as the pixels. Finally, a winner-takes all approach was applied to annotate the 3D points located inside these cuboids with the respective 2D pixel class.

This algorithm was implemented using a script written in Matlab. In this regard, the approach uses the planar nature of building façades as an advantage in performing a 2.5D operation, whereas this feature is sometimes considered one of the reasons behind the failure of more direct 3D deep learning methods in identifying classes within a building façade.

Figure 5. Result of the semantic segmentation on the test data superimposed on the original orthophoto.

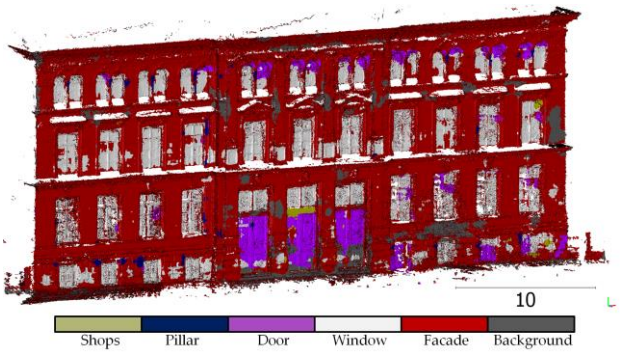

Figure 6. Result of the back-projection to the 3D space: segmented and classified point cloud.

		GROUND TRUTH						
	Class	**Window**	**Door**	**Shops**	**Pillar**	**Facade**	**Background**	***Total***
PREDICTED	**Window**	**2 374 435**	40 542	0	45 088	686 323	8 251	*3 154 639*
	Door	199 202	**370 335**	633	6 119	182 298	11 542	*770 129*
	Shops	9 374	4 348	**14 573**	0	17 770	707	*46 772*
	Pillar	0	2 606	0	**29 838**	88 917	3 599	*124 960*
	Facade	943 476	53 311	2 796	220 726	**14 219 690**	1 099 793	*16 539 792*
	Background	8 977	92 834	234	16	639 034	**256 089**	*997 184*
	Total	*3 535 464*	*563 976*	*18 236*	*301 787*	*15 834 032*	*1 379 981*	***21 633 476***

Table 1. Confusion matrix of the classification results. True positive values are highlighted in blue.

	Window	**Door**	**Shops**	**Pillar**	**Facade**	**Background**
Errors of commission (%)	24.73	51.91	68.84	76.12	14.03	74.32
Errors of omission (%)	32.84	34.33	20.09	90.11	10.20	81.44
Precision/Producer accuracy (%)	75.27	48.09	31.16	23.88	85.97	25.68
Recall/User Accuracy (%)	67.16	65.67	79.91	9.89	89.80	18.56
F1 Score (%)	70.98	55.52	44.83	13.98	87.85	21.55
Intersection over Union (%)	70.92	65.71	82.59	11.66	99.73	22.54

Table 2. Statistics derived from the result of the semantic segmentation.

4. RESULTS AND DISCUSSIONS

The result of the semantic segmentation can be seen in Figure 5. Visually, while the algorithm managed to correctly determine the class of most of the pixels, some problems can also be observed. Indeed, misclassification is seen in the ground and second levels of the building. The texture data for the ground level was plagued by shadows and this may play an important role in explaining these results. As for the errors in the second floor, the quality of the original 3D reconstruction deteriorated as the point of view is farther from the ground. This is due to the terrestrial nature of the original data acqusition, therefore generating holes and textural distortions on higher positions.

To add to these explanations, the orthophoto, unlike the rectified images of the training dataset, is a reconstructed raster. The orthophoto pixel grey values were reconstructed from the projection of input image pixel RGB values into the depth map. This means that in places where the quality of the depth map is low (notably the borders of the building and again, higher storeys) distorted textures may be present. In turn this also influences the quality of the semantic segmentation.

The segmented image as shown in Figure 5 was then back-projected using the aid of the input depth map to the 3D point cloud. The result is a semantically segmented point cloud as displayed by Figure 6.

In order to present a more robust analysis and borrowing from the remote sensing domain, a confusion matrix (Table 1) was prepared in order to illustrate the performance quality of the proposed approach in a quantitative manner. A quick statistical parameter which may be used to represent the quality of multi-class classification is the overall accuracy value, defined as:

$$Overall\ Accuracy = \frac{\sum TruePositives}{Total\ points} \quad (1)$$

Note that the true positive values in Table 1 are highlighted in blue. Using formula (1), an overall accuracy of 0.80 (79.81%) was obtained.

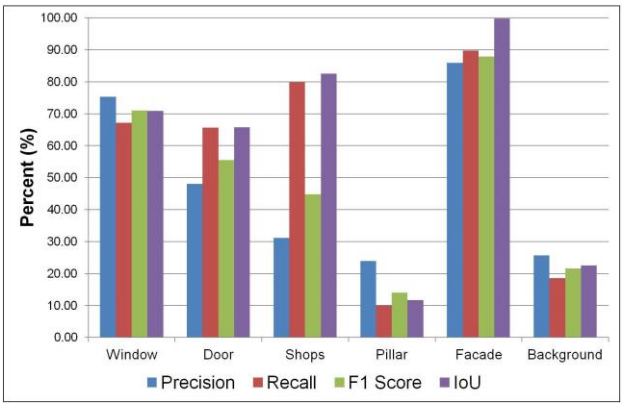

Figure 7. Bar chart illustrating the precision, recall, F1 score, and intersection over union (IoU) of each class.

This overall accuracy value is encouraging; however, some discrepancies between the true positive in each class may be observed in the confusion matrix. In order to analyse this phenomenon in more detail, Table 2 presents other quality metrics commonly used in the field of remote sensing and image segmentation (Heipke et al., 1997, Landes et al., 2012). These statistics were computed as such:

$$Errors\ of\ commission = \frac{FalsePositives}{TruePositives + FalsePositives} \quad (2)$$

$$Errors\ of\ ommission = \frac{FalseNegatives}{TruePositives + FalseNegatives} \quad (3)$$

$$Producer\ accuracy = \frac{TruePositives}{TruePositives + FalsePositives} \quad (4)$$

$$User\ accuracy = \frac{TruePositives}{TruePositives + FalseNegatives} \quad (5)$$

Note that the formula (4) corresponds to the notion of *precision* in ML and DL parlance, while formula (5) corresponds to the notion of *recall*. Furthermore, formulae (2) and (4) represent normalised quantitative value for the incidence of false positives (type I error); as such the sum of their outputs is equal to 1. This is also true for formulae (3) and (5), themselves representing the incidence of false negatives (type II error). Also note that in Table 2, these values are represented in percent in order to synchronise them with ML conventions. In addition, normalised F1 scores were also computed from the outputs of formulae (4) and (5) using the following equation:

$$F1 = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (6)$$

Figure 7 displays the precision, recall and F1 score values in a bar chart for a visual comparison between the different classes. As can be seen from both Table 2 and Figure 7, the developed approach managed to attain good results for some classes, while the performance of some others are underwhelming.

As may be expected, the "facade" class fared the best. This is because this class represent a majority of training data during the DL semantic segmentation. On the other hand, the "pillar" class suffered the most. This is also to be expected for two main reasons. First, the amount of training data concerning pillars was significantly lower because the input database mainly consists of building façades of architectural styles where pillars were not the main feature. Secondly, in the test dataset of the Zoological museum, no free-standing pillar was present. Instead, engaged pillars or faux pillars were present mainly as decorative adornments.

In the context of our particular requirements, the detection of building façade openings i.e. the "window" and "door" classes are particularly interesting. In both classes, the results were quite satisfactory even though improvements may be necessary. The "window" class yielded an F1 score of 70.98% while the "door" class fared worse with a score of 55.52%. Several factors may explain these results. First of all, the confusion between doors and windows in this particular dataset mainly concerns both the second and ground levels of the museum. As has been previously mentioned, the texture data for the ground level was not ideal as a lot of shadow was present. For the second storey, the terrestrial point of view prevented an accurate modelling in the 3D point cloud, which undoubtedly also influenced the generated orthophoto.

Secondly, in some instances the windows were classified either as façades or background. A possible factor amounting to this result is the fact that due to current renovations most of the museum's windows were boarded using tan or pale brown paper sheets, effectively a similar colour to the building façade. Note that in order to validate these hypotheses, more experiments must be conducted to eliminate possible causes. That being said, the results from our developed approach for building openings already shows promising results which significantly outperform some 3D DL techniques for the same task, e.g. results from Pierdicca et al. (2020). However, both DL and ML semantic segmentation develops in a very fast pace and this conclusion must be taken with caution following other recent developments, for example those of Matrone et al. (2020).

5. CONCLUSIONS AND FUTURE WORK

In this paper we presented a novel approach for the semantic segmentation of building façades. Being notoriously difficult to perform in a pure 3D space-based DL approaches, we proposed a two-step process in which the DL semantic segmentation was performed on the building's orthoimage. Helped by a depth map, the results of this segmentation was thereafter back-projected into the 3D space to generate a semantically segmented point cloud. Using the standard metrics for classification, results of this study show an overall accuracy of 79.81%. This is similar to the validation accuracy of 79.7% during the training process, which means that the results may be considered as promising.

Some important remarks can be inferred from the results. As the DL part of the method relies on pixel RGB values, the quality of the orthophoto became an important issue. Due to the terrestrial nature of the data acquisition, several parts of the orthophoto, notably blind spots were distorted by the orthorectification algorithm. Low lighting spots also present more error when the result is consulted visually. A possible solution to improve the quality of the orthophoto would be the addition of images from more favorable angles, either using a drone or other means. The use of other sources of input e.g. MMS and TLS can also be an alternative, as the orthophoto input can be generated from a projection of point cloud into a 2D surface.

Considering these promising results, more experiments and tests are planned. Some of the planned improvements include the tuning of the DL hyperparameters (including choice of pre-trained network), improvement in orthoimage quality, and tests on other sources of point cloud data. Comparisons with other algorithmic approaches as well as 3D-based DL is also planned in the near future. In supporting the overall task of building façade classification, a DL-based semantic photogrammetry workflow is also being considered as an alternative method.

ACKNOWLEDGEMENTS

This research is part of the project Building Indoor/Outdoor Modeling (BIOM) funded by the French National Agency for Research (ANR) with project number ANR-17-CE23-0003. The authors wish to thank Dr. Hélène Macher for her help in providing surveying data for the Zoological museum as well as several related Matlab scripts.

REFERENCES

Bassier, M., Vergauwen, M., Van Genechten, B., 2017. Automated Classification of Heritage Buildings for As-Built BIM using Machine Learning Techniques, in: ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences Vol. IV-2/W2. pp. 25–30.

Chen, L.C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H., 2018. Encoder-decoder with atrous separable convolution for semantic image segmentation. Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics) 11211 LNCS, pp. 833–851.

Grilli, E., Dininno, D., Petrucci, G., Remondino, F., 2018. From 2D to 3D supervised segmentation and classification for cultural heritage applications, in: The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences Vol. XLII-2. pp. 399–406.

Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., Liu, T., Wang, X., Wang, L., Wang, G., Cai, J., Chen, T., 2017. Recent Advances in Convolutional Neural Networks. Pattern Recognition, 77, pp. 354-377.

Hafiz, A.M., Bhat, G.M., 2020. A survey on instance segmentation: state of the art, in: International Journal of Multimedia Information Retrieval 9, 171–189.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 770–778.

Heipke, C., Mayer, H., Wiedemann, C., Jamet, O., 1997. Evaluation of automatic road extraction. Int. Archives of Photogrammetry and Remote Sensing, Vol. 32, pp.47-56.

Heipke, C., Rottensteiner, F., 2020. Deep learning for geometric and semantic tasks in photogrammetry and remote sensing, Geo-spatial Information Science, 23:1, pp. 10-19

Kaushik, R., Kumar, S., 2019. Image Segmentation Using Convolutional Neural Network, in: International Journal of Scientific and Technology Research, 8, 9.

Kirillov, A., He, K., Girshick, R., Rother, C., Dollar, P., 2019. Panoptic segmentation, in: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. pp. 9396–9405.

Landes, T., Boulaassal, H., Grussenmeyer, P., 2012. Quality assessment of geometric façade models reconstructed from TLS Data. The Photogrammetric Record, 27(138): pp. 137–154

Maalek, R., Lichti, D.D., Ruwanpura, J.Y., 2019. Automatic recognition of common structural elements from point clouds for automated progress monitoring and dimensional quality control in reinforced concrete construction. Remote Sensing. 11, 1102.

Malinverni, E.S., Pierdicca, R., Paolanti, M., Martini, M., Morbidoni, C., Matrone, F., Lingua, A., 2019. Deep learning for semantic segmentation of point cloud, in: The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences Vol. XLII-2/W15. pp. 735–742.

Matrone, F., Grilli, E., Martini, M., Paolanti, M., Pierdicca, R., Remondino, F., 2020. Comparing machine and deep learning methods for large 3D heritage semantic segmentation. ISPRS Int. J. Geo-Information 9(9), 535.

Meyer, P., Noblet, V., Mazzar, C., Lallement, A., 2018. Survey on deep learning for radiotherapy, Computers in Biology and Medicine, Elsevier, Vol. 98 pp. 126–146

Murtiyoso, A., Veriandi, M., Suwardhi, D., Soeksmantono, B., Harto, A.B., 2020. Automatic Workflow for Roof Extraction and Generation of 3D CityGML Models from Low-Cost UAV Image-Derived Point Clouds. ISPRS International Journal of Geo-Information 9, 743.

Neusch, T., Grussenmeyer, P., 2003. Remote sensing object-oriented image analysis applied to half-timbered houses. XIXth CIPA International Symposium, Antalya, Turkey, Sept. 30-Oct. 4, 2003. In: The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences Vol. XXXIV, Part 5/C15, pp. 298-301.

Pierdicca, R., Paolanti, M., Matrone, F., Martini, M., Morbidoni, C., Malinverni, E.S., Frontoni, E., Lingua, A.M., 2020. Point Cloud Semantic Segmentation Using a Deep Learning Framework for Cultural Heritage. Remote Sens. 12, 1005.

Qi, C. R., Su, H., Kaichun, M., Guibas, L.J., 2017a. PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation, in: 2017 IEEE Conference on Computer Vision and Pattern Recognition IEEE, Honolulu, HI, pp. 77–85.

Qi, C.R., Yi, L., Su, H., Guibas, L.J., 2017b. PointNet++: Deep hierarchical feature learning on point sets in a metric space. Advances in Neural Information Processing Systems. 2017-December, pp. 5100–5109.

Tang, W., Zou, D., Yang, S., Shi, J., Dan, J., Song, G., 2020. A two-stage approach for automatic liver segmentation with Faster R-CNN and DeepLab, in: Neural Computing and Applications32, pp. 6769–6778.

Tyleček, R., Šára, R., 2013. Spatial pattern templates for recognition of objects with regular structure. Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics) 8142 LNCS, pp. 364–374.

Zolanvari, S.M.I., Ruano, S., Rana, A., Cummins, A., da Silva, R.E., Rahbar, M., Smolic, A., 2019. DublinCity: Annotated LiDAR Point Cloud and its Applications. arXiv preprint arXiv:1909.03613.