Pointnet For The Automatic Classification Of Aerial Point Clouds

Soilán, M.; Lindenbergh, R.; Riveiro, B.; Sánchez-Rodríguez, A.

DOI
10.5194/isprs-annals-IV-2-W5-445-2019

Publication date
2019

Document Version
Final published version

Published in
ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences

Citation (APA)
Soilán, M., Lindenbergh, R., Riveiro, B., & Sánchez-Rodríguez, A. (2019). Pointnet For The Automatic Classification Of Aerial Point Clouds. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *4*(2/W5), 445-452. https://doi.org/10.5194/isprs-annals-IV-2-W5-445-2019

POINTNET FOR THE AUTOMATIC CLASSIFICATION OF AERIAL POINT CLOUDS

M. Soilán [1]*, R. Lindenbergh [2], B. Riveiro[1], A. Sánchez-Rodríguez[1]

[1] Dept. of Materials Engineering, Applied Mechanics and Construction, School of Industrial Engineering, Univeristy of Vigo, 36310, Vigo, Spain – (msoilan, belenriveiro, anasanchez)@uvigo.es
[2] Dept. of Geoscience & Remote Sensing, Faculty of Civil Engineering and Geosciences, Delft University of Technology, 2628 CN Delft, The Netherlands – (R.C.Lindenbergh@tudelft.nl)

Commission II, WG II/10

KEY WORDS: Aerial Laser Scanner, Point Cloud Classification, Deep Learning, Semantic Segmentation.

ABSTRACT:

During the last couple of years, there has been an increased interest to develop new deep learning networks specifically for processing 3D point cloud data. In that context, this work intends to expand the applicability of one of these networks, PointNet, from the semantic segmentation of indoor scenes, to outdoor point clouds acquired with Airborne Laser Scanning (ALS) systems. Our goal is to of assist the classification of future iterations of a national wide dataset such as the *Actueel Hoogtebestand Nederland* (AHN), using a classification model trained with a previous iteration. First, a simple application such as ground classification is proposed in order to prove the capabilities of the proposed deep learning architecture to perform an efficient point-wise classification with aerial point clouds. Then, two different models based on PointNet are defined to classify the most relevant elements in the case study data: Ground, vegetation and buildings. While the model for ground classification performs with a F-score metric above 96%, motivating the second part of the work, the overall accuracy of the remaining models is around 87%, showing consistency across different versions of AHN but with improvable false positive and false negative rates. Therefore, this work concludes that the proposed classification of future AHN iterations is feasible but needs more experimentation.

1. INTRODUCTION

The extraction of semantic information from 3D point cloud data in a reliable manner has been a challenge over the last decade. Point clouds are a useful source of data for a large number of applications, such as 3D modelling, urban and itinerary planning or virtual tourism. But they are also unorganized and unstructured, and therefore basic operations such as point neighbourhood definition are not as trivial, depending on variables as the point density or the number of points within the point cloud unlike other data sources such as 2D images. Some of the applications related to object detection considered building reconstruction from aerial point clouds (Lafarge et al., 2008; Ortner et al., 2007), as well as tree detection and reconstruction (Wang et al., 2008; Xu et al., 2007). An object detection and recognition framework was proposed by Golovinskiy et al. (2009), which consists of an initial point clustering that segments objects in specific locations of interest, and then defines a number of features that gather information of each object in terms of shape and context. Finally, different classifiers such as Support Vector Machines (SVM) or Random Forest are employed to assign a class to each object. This approach has been repeatedly used in literature. Typically, a first differentiation of ground and off-ground points is made, then off-ground objects are clustered and classified, being the last step the main difference between similar works. A common application is the classification of objects related with the road network using Mobile Laser Scanning (MLS) systems. Pu et al. (2011) recognize poles and trees, as well as the shape of traffic signs, using a knowledge based analysis of the geometry of each object. Similarly, Yu et al. (2015) extract light poles along the road network by defining a pairwise 3-D shape context feature which classifies light poles using a threshold over a dissimilarity measure. Detection and recognition of infrastructure related objects have evolved during this decade using both heuristic and machine learning based approaches (Arcos-García et al., 2017; Cheng et al., 2017; Sánchez-Rodríguez et al., 2018; Soilán et al., 2017; Wen et al., 2015), as well as generic object classification approaches (Serna and Marcotegui, 2014; Yang et al., 2015).

During the last couple of years, there has been a considerable increase of works that develop deep learning networks specifically for processing 3D point cloud data. Some of them rely on projecting the point cloud onto 2D images that are classified using 2D Convolutional Neural Networks (CNN) designed for semantic segmentation of images, and then projecting the image back to 3D space. Lawin et al. (2017) project the point cloud in different 2D views and render different properties such as colour and depth from each view. Then, they process each image with a CNN and aggregate individual predictions from all images where a point is visible to assign a label to it. Similarly, Boulch et al. (2017) generate random camera positions as point cloud projection centres, allowing to define RGB and depth composite images. Different existing CNN networks such as SegNet or U-Net are subsequently trained for semantic segmentation of the images, and the result is projected back on the point cloud. These methods alleviate the drawbacks of initial true 3D deep learning approaches, that required a voxelization that led to a decrease of the spatial resolution and a large consumption of memory. However, there are different deep learning approaches that work natively with 3D point clouds. Engelcke et al. (2016) proposed a sparse convolution over a 3D grid space where a feature vector is defined for each grid cell that has at least one point. The sparsity of the convolution helps to avoid the fact that in a 3D convolution across a point cloud, computation time is wasted given that a large number of voxels is empty. Tatarchenko et al. (2017) define a convolutional decoder architecture that generates 3D outputs

* Corresponding author

represented as octrees, scaling efficiently to higher resolutions and improving some of the drawbacks of the 3D deep learning approaches. Another relevant architecture for semantic segmentation, PointNet (Qi et al., 2016) will be explored in this paper and is explained in detail in Section 3.1. In conclusion, deep learning based approaches seem to improve previous methods. For example, some of the work related to the detection and recognition of infrastructure objects improve the previous state of the art (Wen et al., 2019; Yu et al., 2016).

With this context, it is clear that state-of-the-art methods can be applied to automatize point cloud labelling processes, that are typically carried out in a semiautomatic or manual manner, requiring an investment on the process that do not add value to the personnel in charge of the labelling process. Nowadays public administrations develop actions to collect point cloud data from the national territory (e.g. for The Netherlands, Spain, Denmark or Switzerland) that end up being a massive amount of points which, if labelled, can be used as benchmark for a large number of applications and save economic resources invested on the labelling process.

Therefore, the contribution of this work is twofold. First, to prove the capabilities of the PointNet architecture (originally intended for indoor data) for semantic segmentation applications in outdoor point clouds, and second, to evaluate if that architecture can be employed to assist the classification of future versions of a national aerial point cloud database (in the case of this work the database is from The Netherlands), given that there already exists a previous labelled version.

In Section 2, the case study data employed in this work is described. Then, in Section 3, the methodological approach followed for the definition of three different classification models is explained. The evaluation of the classification models and the discussion of the results can be found in Section 4, and finally the conclusions of this work are outlined in Section 5.

2. CASE STUDY DATA

All the data employed to undertake this work has been obtained from the *Actueel Hoogtebestand Nederland* (AHN), which has collected aerial point clouds of The Netherlands. Currently, the latest version of AHN 3D point cloud data is the AHN3 dataset (PDOK, 2018). This dataset is interesting for this work, as it has a decent point density of about 20 points/m^2, and a label assigned to each point, indicating up to five different classes, namely: Ground, vegetation, building, bridge, and water. Furthermore, AHN3 data has several point attributes, from which intensity and return number are utilized in this work. Besides AHN3, in order to assess the performance of the classification models defined in Section 3 with different iterations of AHN, the previous version of these data, AHN2, is also employed. Although the point density is similar, these point clouds do not have measured attributes such as point intensity, nor class labels, facts that are reflected in the chosen classification models in Section 3.2.

The data from AHN3 and AHN2 that is employed in this work are summarized in Table 1. Note that both datasets are divided in smaller point clouds following a rectangular grid, each of them covering a surface of 6.25x5 km^2 and defined with a unique ID. The number of points shown for each section corresponds with the amount of data used for either training or testing, and does not represent the whole AHN section, as detailed in Section 3 of this work.

Figure 1 shows the location of the point clouds in the map. As it can be seen, AHN2 data intentionally overlaps with a section of AHN3 data, in order to compare the results in Section 4. The only criteria that was taken into consideration for selecting training data was the presence of different environments (urban, countryside) in order to train models with as many different geometries as possible.

Dataset	Section ID	#Points·10^6	Usage
AHN3	38FN1	30	Training
	31HZ2	30	Training
	32CN1	25	Test
	37EN2	20	Test
AHN2	37EN2	20	Test

Table 1. Overview of used data from the second (AHN2) and third (AHN3) version of the Dutch national point cloud archive `Actueel Hoogtebestand Nederland` (AHN)

Figure 1. Map of the AHN sections selected as case study. Training data from the surroundings of Utrecht, and test data also from the area of Delft

3. METHODOLOGY

In this work, the capability of the PointNet architecture for the semantic segmentation of aerial and terrestrial point clouds is assessed using the case study data presented in Section 2. First, the main characteristics of this network are outlined. Then, the strategy followed for training the PointNet models is explained in detail, including data preparation and organisation, and also result presentation.

3.1 PointNet architecture

This work employs PointNet, a Deep Learning architecture specifically designed for 3D point clouds (Qi et al., 2016). It considers 3D points as inputs, assuming three main properties: (1) Data is unordered (any permutation of 3D points results in the same set of points); (2) there is interaction among points (neighbourhood relationships are meaningful), and (3) data is invariant to rotations and translations in terms of semantic meaning of each point. With these considerations, the network can be roughly defined with three key modules: (1) A symmetry function that inputs n vectors and outputs another vector invariant to input order, making a permutation invariant model; (2) aggregation of local and global information, by concatenating a global point cloud feature vector with per point features; and (3) joint alignment network, called T-net in the literature, that predicts an affine transformation which makes the network

invariant to transformations such as rotations and translations (rigid transformations).

This architecture is suitable for different applications, such as object classification, part segmentation, or semantic segmentation. In this work, semantic segmentation has a greater interest, as the objective is to offer a point-by-point classification of point cloud data.

This architecture has proven to be efficient for semantic segmentation of indoor point clouds such as the Stanford 3D dataset (Armeni et al., 2016), following a tiling strategy of the point cloud, and randomly selecting a number of points of each tile block at training time, ending up with a 9-dimensional vector representation of each point: xyz coordinates, rgb colour and the normalized xyz coordinates of the point within the tile block (that is, the coordinates of each tile block are normalized to a [0,1] range). However, the performance of this network on outdoor point clouds has not been explored. Doing this is one of the principal interests of this work.

3.2 Classification models

Given the case study data as shown in Section 2 and their properties, three different PointNet based models were defined for semantic segmentation, as summarized in Table 2.

Model	Database	Features	Classes
Model 1	AHN3	*x, y, z, I, R, H*	*2 (Ground, Not Ground)*
Model 2	AHN3 AHN2	*x, y, z*	*3 (Ground, Vegetation, Building)*
Model 3	AHN3	*x, y, z, I, R, H*	*3 (Ground, Vegetation, Building)*

Table 2. Summary of the classification models

3.2.1 Model 1: In order to assess the capabilities of the PointNet architecture for the semantic segmentation of aerial point clouds, a binary classification of the ground was initially proposed. That means, the labels of the training data are modified in such a way that all points which are not classified as ground share the same label.

Some modifications were carried out for the correct implementation of the PointNet architecture with respect to the original network. First, RGB colour information is not included in AHN3, while this point feature was included in previous PointNet setups. Instead, colour was replaced by three features which could be considered as a false colour: (1) Intensity (I), (2) return number (R) and (3) height of the point with respect to the lowest point in a 3x3m neighbourhood (H) (Figure 2). Furthermore, as our model was intended to perform a binary classification, the output layer of the network was modified, defining only two output units.

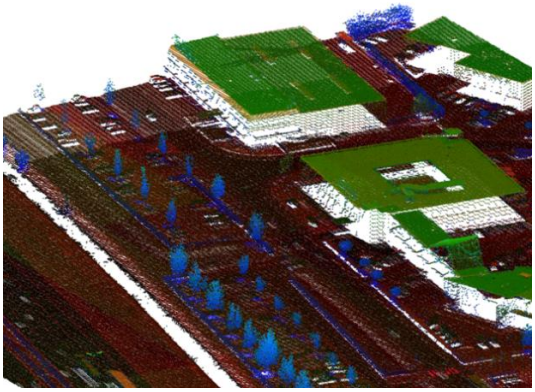

Figure 2. RGB colour has been replaced by three features, namely point intensity, return number, and height with respect the lowest point in a neighbourhood. Visually, it can be seen that the ground has a false colour easily distinguishable from other points

The data organization parameters were also modified, as originally each point cloud was sampled into 1x1m blocks. Given that the density of the aerial point clouds is smaller, and the surface covered much larger, the block surface was set to 10x10m, with a stride of 7.5m, which allows to capture efficiently the local geometric properties of the point cloud in each block. A total of 4096 points are sampled from each block in order to define data batches with a consistent number of points. If the number of points within the block is larger, they are picked randomly. Otherwise, random points are duplicated to get the required number of points per data batch. Hence, the network is fed by a Nx4096x9 array, where N is the number of data batches, and for each point a 9-dimensional feature vector is defined as shown in Equation 1:

$$feat_{model1} = (x, y, z, I, R, H, x_n, y_n, z_n) \quad (1)$$

where (x, y, z) are the point coordinates, (I, R, H) are the previously defined features and (x_n, y_n, z_n) are the normalized coordinates as defined in Section 3.1.

With these considerations, the network was trained with data from the 38FN1 section of AHN3, sampling around 30 million points by tiling the whole section using a 25x25 square grid and randomly selecting one tile. The parameters used for training scarcely differ from the default, training for 50 epochs, mini batch size of 24, using Adam (Kingma and Ba, 2015) as optimizer, and regularization with an initial learning rate of 0.001, with decay following a staircase function each 300 thousand training samples.

The results obtained for the test set, that are shown in Section 4.1, prove that PointNet is capable of performing semantic segmentation tasks in AHN point clouds; hence the proposal of assisting the classification of future AHN iterations with a multiclass classification model.

3.2.2 Model 2: Considering that AHN3 is currently the latest iteration within AHN, AHN2 dataset was chosen to evaluate the capability of the PointNet architecture for classifying point clouds from different versions of the same data. That is, the objective of this model is to perform efficiently in AHN2 data while being trained with AHN3 point clouds.

This implied some issues: First, AHN2 point clouds do not have a classification field and therefore to generate a ground truth for evaluation is not straightforward. In order to solve this, a nearest neighbour search was performed between the AHN2 data and overlapping data from AHN3. Then, the label of each point in AHN3 was assigned to the closest point in AHN2, defining a ground truth for evaluation. Here, it is important to have in consideration a rough estimate of the proportion of mislabelled points in this process, which is directly related to changes (new buildings, different distribution of vegetation, etc.) in the time between the collection of both datasets. For that purpose, the distances between nearest neighbours in AHN2 and AHN3, $\boldsymbol{d}_{2\to3}$ are obtained, and a rough estimate of mislabelled points is extracted from the proportion of points such that $\boldsymbol{d}_{2\to3}$ is larger than the average μ plus the standard deviation σ of the distribution of distances (Figure 3a-b).

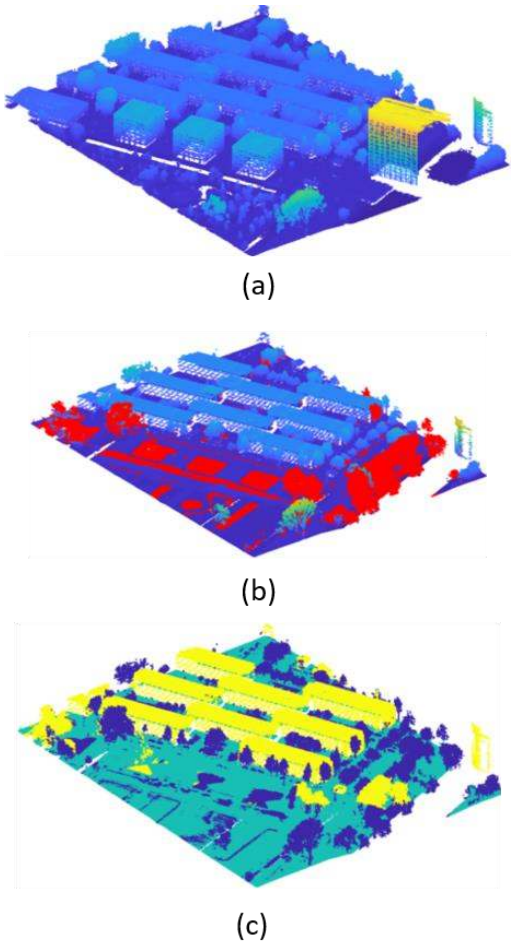

Figure 3. In order to get a rough estimate of mislabelled points in AHN2 after assigning the closest label from AHN3 to each point, the assignment distance is employed. (a) AHN3 point cloud. (b) AHN2 point cloud, points in red represent areas subjected to changes between both scans that are therefore likely to be wrongly labelled. (c) point cloud labelled with three classes: Ground (cyan), vegetation (dark blue) and buildings (yellow)

Regarding the labels that are used for training the network, Model 2 considers three different classes, namely: (1) Ground, (2) vegetation, and (3) building (Figure 3c). Note that both the bridge and water classes are not considered here, as they are excessively unbalanced in terms of number of points. While points classified as bridge are assigned to the ground class, those classified as water are removed from the training set.

Another issue arises from the fact that AHN2 point clouds have less attributes than their AHN3 counterpart (i.e. it has no intensity). For that reason, this model could not take advantage of the features that have been employed in Model 1. Therefore, this model is defined as an end-to-end strategy that only considers the geometry, that is, using only the point coordinates as features. This way, and following the same data organization guidelines defined for Model 1 in Section 3.2.1, the network is fed with points whose feature is simply:

$$feat_{model2} = (x, y, z, x_n, y_n, z_n) \tag{2}$$

The network was trained with data from section 31HZ2 as introduced in Section 2, and using a similar number of points as for Model 1, which were considered enough training samples given its results. Training parameters were also similar, only the number of training epochs was increased from 50 to 70.

This classification model was evaluated both for AHN3 and AHN2 data, using an overlapping area in order to offer a better comparison of the results, which can be seen in Section 4.2.

3.2.3 Model 3: Model 2 has been trained using only the coordinates (x, y, z) as feature, but it is also interesting to assess the impact of training the same network with more features, as it can be assumed that future iterations of AHN dataset will present at least the same point cloud properties as AHN3. For that reason, Model 3 has been defined for the same classification problem as Model 2, that is, a 3-class classification of ground, vegetation and buildings. However, three new features were added following the reference of Model 1. That is, the point feature is a 9-dimensional vector analogous to $feat_{model1}$:

$$feat_{model3} = (x, y, z, I, R, H, x_n, y_n, z_n) \tag{3}$$

Regarding the training data, in order to stablish a comparison between Model 2 and Model 3 evaluations, the same data from AHN3 section 31HZ2 was chosen. Also, data organization and network parameters follow the same guidelines as for Model 2.

3.3 Presentation of the results

Once the different models are trained, they are evaluated using the test datasets defined in Section 2. The model evaluation process outputs a NxM array, where N is the number of points whose prediction is given by the model, and $M = 5 + K$, where K is the number of classes predicted by the model. For each point, a vector $output = (x, y, z, pred, gt, p_0, \dots, p_k)$ is obtained, which contains the point coordinates (x, y, z), the point label as predicted by the network (*pred)*, the point label of the ground truth (*gt)* and the probabilities of each class for the given point $(p_0, \dots, p_k)$. Using the outputs $pred$ and gt it is straightforward to define true positives, false positives and false negatives, obtaining different metrics chosen for evaluation: Precision, Recall and F-score (Equations 4-6). Results shown in Section 4 can be produced with this data.

$$Precision = \frac{TP}{TP + FP} \tag{4}$$

$$Recall = \frac{TP}{TP + FN} \tag{5}$$

$$F_{score} = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \quad (6)$$

where TP = number of true positives
FP = number of false positives
FN = number of false negatives

4. RESULTS AND DISCUSSION

In this section, the evaluation results for the three classification models based on the PointNet architecture are first presented and discussed one by one, followed by a global discussion to put the results in perspective.

In order to train the different networks, resources from the computing infrastructure FinisTerrae II from the Supercomputing Centre of Galicia (CESGA) were employed. The GPUs available for the training processes were NVIDIA Tesla K80, and training times ranged from 8 to 12 hours depending on the model complexity and the number of training epochs.

4.1 Model 1 results

The objective of the first model was to perform a conceptually simple task on AHN point clouds, which is ground classification, in order to assess the capability of the PointNet architecture to carry out more complex applications. The test data employed to evaluate this model consists of around 25 million points from the 32CN1 tile of AHN3 as shown in Section 2. In Table 3 the results are summarized. Results are also compared with (Rizaldy et al., 2018) where a fully convolutional network (generating 2D images from the point cloud in a first place) is employed for the same ground classification application in AHN3 data.

	Precision	Recall	Fscore
This work	0.955	0.970	0.962
(Rizaldy et al., 2018)	0.849	0.959	0.901

Table 3. Model 1 results. Our PointNet results improve upon previous results from (Rizaldy et al., 2018)

An example of qualitative results can be seen in Figure 4. The main conclusion extracted from these results is that the trained model performs correctly, as it defines the ground segment with high accuracy with only few erroneous predictions. These results are good enough across a large surface of AHN3 to develop more complex classification models with this architecture.

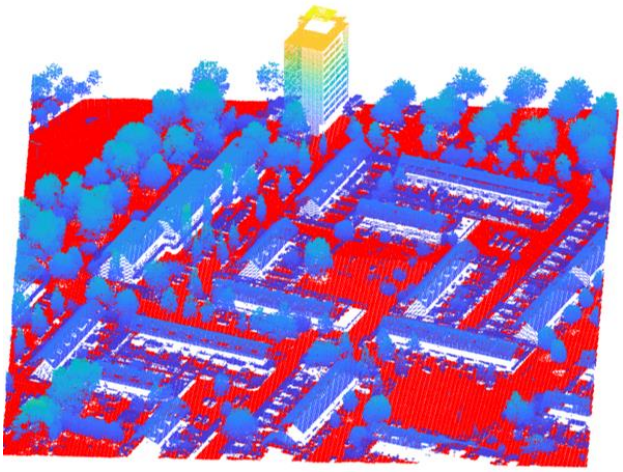

Figure 4. Ground classification using Model 1. Points predicted as ground are coloured in red

4.2 Model 2 results

The second model defines a three-class classification using AHN3 data for training. In order to assess the capability of the trained model to label data from different versions of AHN, data from both AHN3 and AHN2 are used for the evaluation of the model.

The test data comprises around 20 million points from the 37EN2 section of AHN3 and AHN2. Data from both datasets overlaps, which allows to define a ground truth in AHN2 data as described in Section 3.2.2. The results of the prediction for both datasets are shown in Tables 4-5. Here, a confusion matrix is presented for each case, showing the false positive and false negative rate, together with the overall accuracy defined as the ratio between points that are correctly predicted by the classification model and the total number of points within the test set. Qualitative results can be seen in Figure 5.

GT/ predict	Ground	Vegetation	Building	FN rate
Ground	**10 233 948**	225 557	138 446	3.43%
Vegetation	322 025	**8 677 499**	124 554	4.89%
Building	340 387	1 635 044	**1 679 099**	54.05%
FP rate	6.08%	17.66%	13.54%	
Accuracy	**87.77%**			

Table 4. Model 2 results for the AHN3 test set. Note the confusion between building and vegetation points

GT/ predict	Ground	Vegetation	Building	FN rate
Ground	**8 727 286**	418 813	138 288	6.00%
Vegetation	293 896	**3 005 796**	99 892	11.58%
Building	328 280	1 034 028	**1 389 390**	49.51%
FP rate	6.65%	32.58%	14.63%	
Accuracy	**84.98%**			

Table 5. Model 2 results for the AHN2 test set. For AHN2 similar results are obtained as for AHN3 (compare Table 4), indicating that it is feasible to automatically classify future AHN releases using our approach

Some conclusions can be extracted from these results. First, in terms of network performance, the predictions on AHN3 and AHN2 datasets are fairly similar. Although the performance on AHN2 is slightly less than on AHN3, it is important to recall that generating the ground truth data for the former has a small mislabelling rate (around 2-3%) as defined in Section 3.2.2. In conclusion, classifying future iterations of AHN with a model trained in a previous one seems feasible, as far as the model performs correctly in terms of accuracy. Considering the performance metrics in Tables 4-5, high confusion between vegetation and building points is observed, which is illustrated in Figure 5, while the ground class is classified with similar accuracies as for Model 1.

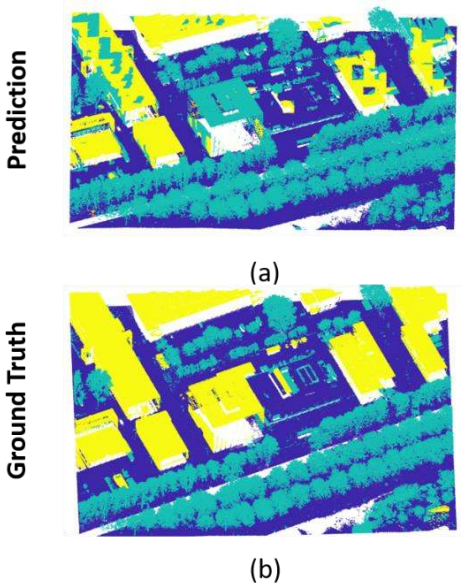

Figure 5. Classification results using Model 2 on a subset of the AHN2 test data. (a) Prediction (ground, vegetation and buildings are coloured as blue, green and yellow respectively). (b) Ground truth. Many roof points are classified as vegetation

4.3 Model 3 results

Finally, a third PointNet model was trained with the same objective as Model 2, a three-class classification, but instead of relying only on the geometry, more features were considered for the training process.

In order to compare the results obtained from the evaluation of this model, the same test set as for Model 2 was used, that is, data from 37EN2 section of AHN3 dataset. In Table 6, a confusion matrix with the results is shown. It can be seen that the results are practically the same as those in Table 4 for Model 2. This implies that adding new features does not seem to have an impact at all on the network performance, and it relies purely on the geometry of the point cloud with its coordinates (x, y, z). A visual example of these results in a particular case where the confusion between vegetation and building classes is noteworthy is shown in Figure 6. Qualitatively, it was noted that the network tends to fail classifying high buildings and their façades, assigning high probabilities to the vegetation class to those points.

GT/ predict	Ground	Vegetation	Building	FN rate
Ground	**10 329 433**	93 115	175 516	2.53%
Vegetation	400 765	**8 607 694**	115 521	5.66%
Building	449 368	1 637 483	**1 567 664**	57.10%
FP rate	7.60%	16.74%	15.66%	
Accuracy	**87.37%**			

Table 6. Model 3 results.

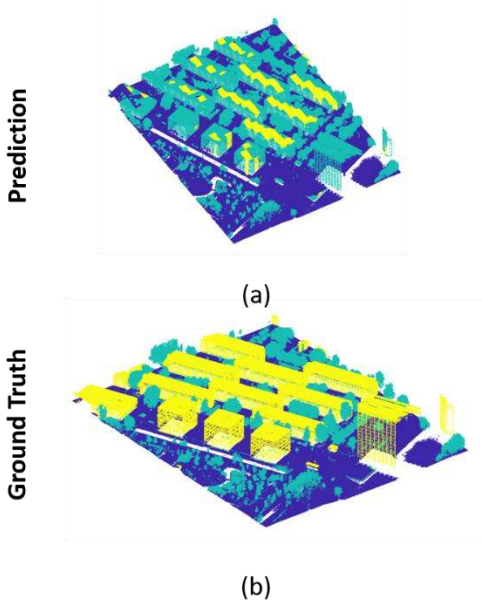

Figure 6. Classification results using Model 3 on a subset of the AHN3 data with a high confusion between vegetation and buildings. (a) Prediction (ground, vegetation and buildings are coloured as blue, green and yellow respectively). (b) Ground truth. Again many roof points are classified as high vegetation

4.4 Global discussion

This work proposed two main contributions in Section 1: To prove the capabilities of the PointNet architecture for semantic segmentation applications in outdoor point clouds, and to evaluate the specific application of classifying point clouds for future versions of the AHN dataset. Regarding the first one, a simple yet demonstrative application (ground classification) was considered, showing high classification accuracies and motivating the second contribution. However, the three-class classification proposed with models 2 and 3 showed that there is still room for improvement. While the ground segment of the point clouds is classified with decent performance metrics, there is a high confusion rate between the two remaining classes, vegetation and buildings (note that the accuracy values in Tables 4-6 are biased by the large number of ground points). This same problem remains even when features that may help to discriminate between both classes such as the return number and the point intensity are added to the feature vector that is fed to the classification network. In order to solve that, different considerations can be made, such as proposing a hierarchical classification that in a first place classifies the ground, and in a second classification, using different features, distinguishes between buildings and vegetation. Besides the conclusions that have been extracted from the quantitative values of the performance metrics, it can also be concluded, from the

compared results of AHN2 and AHN3 datasets, that is feasible to automatically assist the labelling of future iterations of national-wide point cloud databases.

5. CONCLUSIONS

In this work, an already existing architecture for the semantic classification of point clouds, PointNet, which was initially developed for different applications including semantic segmentation of indoor scenes, was employed in order to assess its suitability for the classification of aerial point clouds from the AHN dataset, an aerial point cloud that covers The Netherlands. For that purpose, three different models based on the same architecture were proposed. The first one performed a binary ground classification, and proved that PointNet is suitable for applications in outdoor point clouds after modifications in some network parameters and in the data organization strategy, obtaining an F-score above 96% for the classification of the ground. This good result motivates the subsequent part of this work, where two different models are trained for classifying ground, vegetation and buildings, the three main elements that are labelled in AHN3. Although results across different versions of AHN are similar and therefore the classification of future iterations of the database with a previously trained model seems feasible, the results are not as positive as for the case of ground classification, with a high confusion between vegetation and building classes in both models.

As future work, a deeper understanding of the network will be required in order to determine if the results can be improved with no significant changes at the training stage, or it is necessary to find different approaches for this semantic segmentation applications, such as PointNet++ network, which takes into account multi-scale context. In essence, more experimentation is needed in order to know if a point-wise semantic segmentation is achievable or recommendable over voxel or image based methods.

ACKNOWLEDGEMENTS

This work has been partially supported by the Spanish Ministry of Science, Innovation and Universities through Human Resources program FPI (Grant BES-2014-067736) and its mobility program (EEBB-I-18-12848). This work has used computational resources from the Supercomputing Centre of Galicia (CESGA).
This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 769255. This document reflects only the author's view and the Agency is not responsible for any use that may be made of the information it contains.

REFERENCES

Arcos-García, Á., Soilán, M., Álvarez-García, J.A., Riveiro, B., 2017. Exploiting synergies of mobile mapping sensors and deep learning for traffic sign recognition systems. *Expert Syst. Appl.* 89, 286-295.

Armeni, I., Sener, O., Zamir, A.R., Jiang, H., Brilakis, I., Fischer, M., Savarese, S., 2016. 3D Semantic Parsing of Large-Scale Indoor Spaces. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 1534–1543. doi:10.1109/CVPR.2016.170

Boulch, A., Saux, B. Le, Audebert, N., 2017. Unstructured Point Cloud Semantic Labeling Using Deep Segmentation Networks. *Eurographics Workshop on 3D Object Retrieval*. doi:10.2312/3dor.20171047

Cheng, M., Zhang, H., Wang, C., Li, J., 2017. Extraction and Classification of Road Markings Using Mobile Laser Scanning Point Clouds. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 10, 1182–1196. doi:10.1109/JSTARS.2016.2606507

Engelcke, M., Rao, D., Wang, D.Z., Tong, C.H., Posner, I., 2017. Vote3Deep: Fast Object Detection in 3D Point Clouds Using Efficient Convolutional Neural Networks. *2017 IEEE International Conference on Robotics and Automation (ICRA). 1355-1361*

Golovinskiy, A., Kim, V.G., Funkhouser, T., 2009. Shape-based recognition of 3D point clouds in urban environments. *IEEE 12th International Conference on Computer Vision*. doi:10.1109/ICCV.2009.5459471

Kingma, D.P., Ba, J.L., 2015. Adam: a Method for Stochastic Optimization. *International Conference on Learning Representations* 1–15. doi:http://doi.acm.org.ezproxy.lib.ucf.edu/10.1145/1830483.1830503

Lafarge, F., Descombes, X., Zerubia, J., Pierrot-Deseilligny, M., 2008. Building reconstruction from a single DEM. 2*6th IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. doi:10.1109/CVPR.2008.4587778

Lawin, F.J., Danelljan, M., Tosteberg, P., Bhat, G., Khan, F.S., Felsberg, M., 2017. Deep projective 3D semantic segmentation. *Computer Analysis of Images and Patterns (CAIP)* 95–107. doi:10.1007/978-3-319-64689-3_8

Ortner, M., Descombes, X., Zerubia, J., 2007. Building outline extraction from digital elevation models using marked point processes. *Int. J. Comput. Vis.* 72, 107-132 doi:10.1007/s11263-005-5033-7

PDOK, https://www.pdok.nl/nl/ahn3-downloads (Last visited 16/01/19).

Pu, S., Rutzinger, M., Vosselman, G., Oude Elberink, S., 2011. Recognizing basic structures from mobile laser scanning data for road inventory studies. *ISPRS J. Photogramm. Remote Sens.* 66, S28–S39. doi:10.1016/j.isprsjprs.2011.08.006

Qi, C.R., Su, H., Mo, K., Guibas, L.J., 2017. PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* 77-85. doi:10.1109/3DV.2016.68

Rizaldy, A., Persello, C., Gevaert, C.M., Elberink, S.J.O., 2018. Fully Convolutional Networks for Ground Classification from LIDAR Point Clouds. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* IV, 4–7.

Sánchez-Rodríguez, A., Riveiro, B., Soilán, M., González-deSantos, L.M., 2018. Automated detection and decomposition of railway tunnels from Mobile Laser Scanning Datasets. *Autom. Constr.* 96, 171–179. doi:10.1016/j.autcon.2018.09.014

Serna, A., Marcotegui, B., 2014. Detection, segmentation and classification of 3D urban objects using mathematical

morphology and supervised learning. *ISPRS J. Photogramm. Remote Sens*. 93, 243–255. doi:10.1016/j.isprsjprs.2014.03.015

Soilán, M., Riveiro, B., Martínez-Sánchez, J., Arias, P., 2017. Segmentation and classification of road markings using MLS data. *ISPRS J. Photogramm. Remote Sens.* 123, 94–103. doi:10.1016/j.isprsjprs.2016.11.011

Tatarchenko, M., Dosovitskiy, A., Brox, T., 2017. Octree Generating Networks: Efficient Convolutional Architectures for High-resolution 3D Outputs. *2017 International Conference on Computer Vision, ICCV*.

Wang, Y., Weinacker, H., Koch, B., 2008. A Lidar point cloud based procedure for vertical canopy structure analysis and 3D single tree modelling in forest. *Sensors* 8, 3938-3951. doi:10.3390/s8063938

Wen, C., Li, J., Member, S., Luo, H., Yu, Y., Cai, Z., Wang, H., Wang, C., 2015. Spatial-Related Traffic Sign Inspection for Inventory Purposes Using Mobile Laser Scanning Data. *IEEE Transactions on Intelligent Transportation Systems*. 17, 27–37. doi:10.1109/TITS.2015.2418214

Wen, C., Sun, X., Li, J., Wang, C., Guo, Y., Habib, A., 2019. A deep learning framework for road marking extraction, classification and completion from mobile laser scanning point clouds. *ISPRS J. Photogramm. Remote Sens.* 147, 178–192. doi:10.1016/j.isprsjprs.2018.10.007

Xu, H., Gossett, N., Chen, B., 2007. Knowledge and heuristic-based modeling of laser-scanned trees. *ACM Transactions on Graphics* doi:10.1145/1289603.1289610

Yang, B., Dong, Z., Zhao, G., Dai, W., 2015. Hierarchical extraction of urban objects from mobile laser scanning data. *ISPRS J. Photogramm. Remote Sens.* 99, 45–57. doi:10.1016/j.isprsjprs.2014.10.005

Yu, Y., Li, J., Guan, H., Wang, C., Yu, J., 2015. Semiautomated Extraction of Street Light Poles From Mobile LiDAR Point-Clouds. *IEEE Trans. Geosci. Remote Sens*. 53, 1374–1386. doi:10.1109/TGRS.2014.2338915

Yu, Y., Li, J., Wen, C., Guan, H., Luo, H., Wang, C., 2016. Bag-of-visual-phrases and hierarchical deep models for traffic sign detection and recognition in mobile laser scanning data. *ISPRS J. Photogramm. Remote Sens*. 113, 106–123. doi:10.1016/j.isprsjprs.2016.01.005