SEMANTIC LABELING OF ALS POINT CLOUDS FOR TREE SPECIES MAPPING USING THE DEEP NEURAL NETWORK POINTNET++

S. Briechle[1], P. Krzystek[1], G. Vosselman[2]

[1] Munich University of Applied Sciences, Munich, Germany - (sebastian.briechle, peter.krzystek)@hm.edu
[2] Faculty of Geo-Information Science and Earth Observation (ITC), University of Twente, Enschede,
the Netherlands - george.vosselman@utwente.nl

KEY WORDS: semantic labeling, ALS point clouds, tree species mapping, deep neural network, PointNet++

ABSTRACT:

Most methods for the mapping of tree species are based on the segmentation of single trees that are subsequently classified using a set of hand-crafted features and an appropriate classifier. The classification accuracy for coniferous and deciduous trees just using airborne laser scanning (ALS) data is only around 90% in case the geometric information of the point cloud is used. As deep neural networks (DNNs) have the ability to adaptively learn features from the underlying data, they have outperformed classic machine learning (ML) approaches on well-known benchmark datasets provided by the robotics, computer vision and remote sensing community. Though, tree species classification using deep learning (DL) procedures has been of minor research interest so far. Some studies have been conducted based on an extensive prior generation of images or voxels from the 3D raw data. Since innovative DNNs directly operate on irregular and unordered 3D point clouds on a large scale, the objective of this study is to exemplarily use PointNet++ for the semantic labeling of ALS point clouds to map deciduous and coniferous trees. The dataset for our experiments consists of ALS data from the Bavarian Forest National Park (366 trees/ha), only including spruces (coniferous) and beeches (deciduous). First, the training data were generated automatically using a classic feature-based Random Forest (RF) approach classifying coniferous trees (precision = 93%, recall = 80%) and deciduous trees (precision = 82%, recall = 92%). Second, PointNet++ was trained and subsequently evaluated using 80 randomly chosen test batches à 400 m^2. The achieved per-point classification results after 163 training epochs for coniferous trees (precision = 90%, recall = 79%) and deciduous trees (precision = 81%, recall = 91%) are fairly high considering that only the geometry was included. Nevertheless, the classification results using PointNet++ are slightly lower than those of the baseline method using a RF classifier. Errors in the training data and occurring edge effects limited a better performance. Our first results demonstrate that the architecture of the 3D DNN PointNet++ can successfully be adapted to the semantic labeling of large ALS point clouds to map deciduous and coniferous trees. Future work will focus on the integration of additional features like i.e. the laser intensity, the surface normals and multispectral features into the DNN. Thus, a further improvement of the accuracy of the proposed approach is to be expected. Furthermore, the classification of numerous individual tree species based on pre-segmented single trees should be investigated.

1. INTRODUCTION

1.1 2D and 3D DNNs

Recently, DNNs have gained huge interest as a segmentation and classification method for 2D and 3D data. Examples for well-known deep convolutional neural networks (CNNs) are VGG-16 (Simonyan et al., 2014), ResNet-50 (He et al., 2016) and Mask R-CNN (He et al., 2017). In the past, benchmark datasets have been published to verify and to compare the performance of neural networks. For 2D datasets, very popular benchmarks are the MNIST database (LeCun et al., 1998), the CIFAR-10 dataset (Krizhevsky et al., 2009) and the ImageNet dataset (Deng et al., 2009). In the remote sensing community, state-of-the-art DL methods have been modified for various use-cases: Gevaert et al. (2018) adjusted a Fully Convolutional Network to the application of Digital Terrain Model (DTM) extraction in challenging areas. The method includes an automatic labeling strategy and outperformed two reference DTM extraction algorithms. Vetrivel et al. (2018) successfully detected severe building damages by combining CNN features from oblique aerial images and 3D features from dense photogrammetric point clouds. Since sensors capable of generating 3D data (i.e. stereo camera systems, LiDAR systems) have gained more and more attention and are now widespread in numerous technical fields, the semantic labeling and classification of 3D irregular and unordered point clouds using DNNs is of major research interest. Tchapmi et al. (2017) introduced a framework called SegCloud to obtain semantic scene labeling on point-level using a 3D fully CNN. Based on a voxelization of the 3D point cloud, their approach was evaluated on indoor and outdoor datasets (i.e. KITTI (Geiger et al., 2013)) and a performance comparable or superior to the state-of-the-art was achieved. Zhou et al. (2018) presented the neural network VoxelNET to detect objects (i.e. pedestrians, cyclists) in 3D point clouds based on the encoding of point clouds into equally spaced 3D voxels. Zhao et al. (2018) classified ALS point clouds via deep features learned by a multi-scale CNN. The method creates a group of multi-scale contextual images for each 3D point and is ranked first on the ISPRS benchmark dataset (ISPRS, 2019). All the mentioned 3D approaches transform the irregular 3D data into regular 3D voxel grids or accumulations of 2D images to advantageously utilize neural networks. In contrast to that, algorithms have been developed that directly use the original dataset in a set of sequenced layers to find a best mapping between the input data and the target predictions. These point-based DNNs directly operate on the point cloud without the need for a prior rasterization or voxelization (Figure 1). Qi et al. (2016) developed a highly efficient and effective type of neural network (PointNet) showing i.e. a high performance on the

https://doi.org/10.5194/isprs-archives-XLII-2-W13-951-2019

shape classification benchmarks ShapeNet (Chang et al., 2015) and ModelNet40 (Wu et al., 2015). The offered applications include object classification, part segmentation and semantic labeling. Since PointNet showed limiting ability concerning the recognition of fine-grained patterns and the generalizability to complex scenes, Qi et al. (2017) introduced an enhanced version called PointNet++. This hierarchical neural network recursively applies PointNet and learns local features from multiple contextual scales. It enables an even more accurate classification of single objects as well as the semantic labeling of large-scale point clouds. PointNet++ outperforms PointNet especially on point sets with varying densities like ScanNet (Dai et al., 2017). For object classification, PointNet++ reaches a classification accuracy of 90.7% on the ModelNet40 dataset (40 object categories) and 84.5% on the ScanNet dataset (20 object categories). As robotics and applications like virtual reality and autonomous driving boost the interest in 3D data, innovative approaches for the semantic labeling and the classification of 3D point clouds are being published high-frequently. Landrieu et al. (2018) proposed a DL-based framework for the semantic segmentation of large-scale point clouds and set a new state-of-the-art for outdoor LiDAR scans (i.e. S3DIS (Armeni et al., 2016) and SEMANTIC3D.NET (Hackel et al., 2017)). After efficiently pre-organizing 3D point clouds into geometrically homogeneous elements called superpoint graphs (SPG), a graph convolutional network manages to learn contextual relationships between object parts. In the same year, Li et al. (2018) presented the neural network PointCNN for feature learning from 3D point clouds by generalizing typical CNNs and achieved on par or better performance on multiple challenging 2D (i.e. MNIST, CIFAR-10) and 3D (i.e. S3DIS) benchmark datasets and tasks.

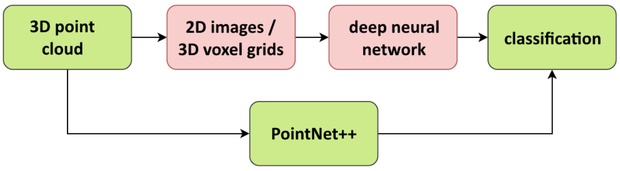

Figure 1. Basic principle of 3D DNNs like PointNet++, operating directly on 3D point clouds without a prior transformation into 2D images or 3D voxel grids

1.2 3D vegetation mapping

Many methods for tree species classification based on segmented single trees use a set of hand-crafted features in combination with an appropriate classifier like RF, Support Vector Machine or logistic regression (Fassnacht et al., 2016). Using only the geometric information of the ALS point cloud, the classification accuracy for coniferous and deciduous trees is around 90%. By extending the feature set with the laser intensity, the accuracy increases to around 95% (Reitberger et al., 2009). DL methods have the ability to automatically learn features and mostly generate more accurate classification results than classic ML approaches using hand-crafted features (see section 1.1). Yet, tree species classification using neural networks has been of minor research interest. Presumably, one reason is the lack of large training datasets. Just recently, Hamraz et al. (2018) use a CNN along with 2D images generated from ALS point clouds to classify coniferous and deciduous trees with 92% and 86% accuracy, respectively. So far, the direct usage of 3D data in DNNs for 3D vegetation mapping is more uncommon.

In this paper, we demonstrate that the architecture of PointNet++ can successfully be adapted to the semantic labeling of large ALS point clouds to map the two tree species spruce and beech.

2. MATERIAL

Airborne full waveform data were acquired in June 2017 (leaf-on condition) using a Riegl LMS-Q 680i instrument which was carried by a plane at a flying altitude of 550 m. The resulting point density was at average 54 points/m^2. The mission area is located in the Bavarian Forest National Park where mainly spruces and beeches are present (95%). Single trees were segmented via the well-known normalized cut (Ncut) segmentation (Reitberger et al., 2009) for the entire area of the National Park (Figure 2). As a baseline method, a RF classifier was trained with 918 manually labeled reference trees (380 coniferous, 538 deciduous) using only geometric features (height dependent and density dependent features, crown shape) to classify the segments with respect to the tree species. Next, the classifier was evaluated using a test dataset comprising 529 trees (293 coniferous, 236 deciduous). For this standard method, the classification results for coniferous trees (precision = 93%, recall = 80%) and deciduous trees (precision = 82%, recall = 92%) were as expected fairly good. Finally, the classifier was used to predict the tree species of all single tree segments. Compared to classic ML approaches like RF, deep learning models require extremely large training datasets in order to capture the essential features in a multilayer structure (Ioannidou et al., 2017). Hence, a study area of 5270 m x 500 m (2.64 km^2) was extracted from the park area comprising approximately 97000 tree segments (48.5% coniferous, 51.5% deciduous) with around 1500 points per tree and a tree density of 366 trees/ha. This dataset was used to train PointNet++ containing 143.3 million points labeled by the classified tree segments and was fully balanced with respect to the two object categories. Potentially misclassified tree segments were not removed by visual inspection.

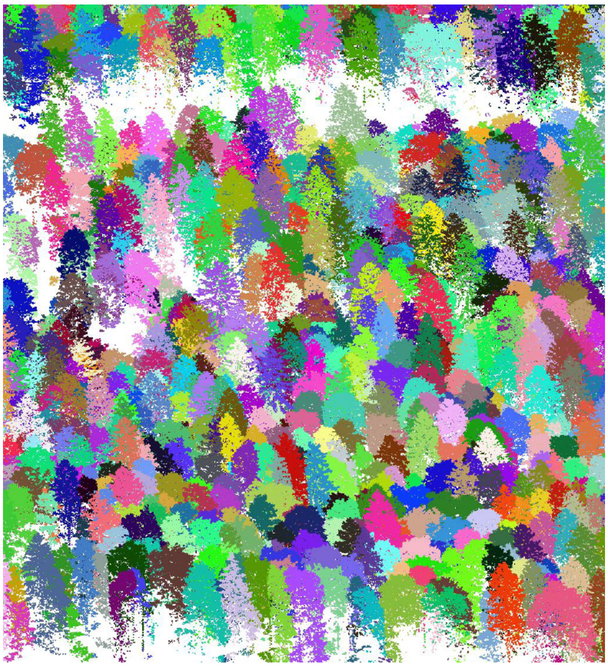

Figure 2. Exemplary single tree segments resulting from the Ncut segmentation; random color rendering

3. METHODOLOGY

The neural network PointNet++ (Qi et al., 2017) operates on unordered 3D data without initially generating images or voxels from the point clouds and calculates per-point scores. It includes a hierarchical feature learning technique as well as special layers that are able to aggregate multi-scale information according to local point densities. For our task, the decisive hyperparameters of the semantic segmentation implementation of PointNet++ (Qi et al., 2019) were adjusted to get a well-performing network (Table 1). Since the dataset comprises two class labels, the activation function of the network was changed from ”softmax” to ”sigmoid”. Next, the training dataset was divided into cubic blocks with an edge length of 60 m (Figure 3). In each training epoch (batch size = 16), smaller batches of 20 m x 20 m x 60 m were extracted and preprocessed including zero centering. Basically, zero centering defines the origin of a local coordinate system as the center of gravity of the selected batches by subtracting the mean X, mean Y, and mean Z values from the absolute coordinates. After each training epoch, the network based on the updated weights was evaluated on randomly chosen test batches. To estimate the performance of the network, standard metrics (precision, recall) were calculated on a single point scale. As graphics processing unit (GPU) we utilized a ”NVIDIA Titan V” (NVIDIA Corporation, 2019).

Hyperparameter	Value	Declaration
BATCH_SIZE	16	Number of batches per epoch.
NUM_POINT	8192	Number of points per batch.
NUM_CLASSES	2	Number of object categories.
MAX_EPOCH	200	Number of training epochs.
BASE_LR	0.001	Initial learning rate.
OPTIMIZER	”adam”	Optimization algorithm.
MOMENTUM	0.9	Momentum value for stochastic gradient descent.
DECAY_STEP	200000	Increment for the reduction of the learning rate.
DECAY_RATE	0.7	Decay rate for the learning rate.
MAX_DROPOUT	0.875	Maximal dropout rate.
CUBE_DIM	60	Edge length of the cubic training blocks in [m].

Table 1. Hyperparameters and default / optimized values

		prediction		
		decid.	conif.	recall
reference	decid.	294623	30516	0.91
	conif.	68707	261514	0.79
	precision	0.81	0.90	
	OA	0.85		
	kappa	0.70		

Table 2. Classification result (epoch 163) for applying the trained neural network on randomly chosen test data; OA = overall accuracy, decid. = deciduous, conif. = coniferous

4. RESULTS AND DISCUSSION

The neural network was trained using 8192 points per batch. Hence, the number of training points per epoch added up to 131072. Assuming approximately 1500 points per tree, this equates to around 87 tree segments per training epoch. After each epoch, the weights of the network were updated. Subsequently, the performance of the network was evaluated on 80 randomly chosen test batches (655360 points $\widehat{=}$ 437 trees $\widehat{=}$ 1.19 ha). After 163 of 200 training epochs (21.4 million points $\widehat{=}$ 14243 trees $\widehat{=}$ 38.9 ha), the classification result for coniferous trees (precision = 90%, recall = 79%) and deciduous trees (precision = 81%, recall = 91%) reached its maximum. Clearly, some experiments were needed to find an optimized set of hyperparameters. The point-based results (Table 2) are fully comparable to the classification results provided by the standard method based on a single tree segmentation and a tree species classification with RF. Appropriate features were automatically generated from the point cloud and are able to discriminate the complex 3D tree structure of both tree species. Nevertheless, it needs to be pointed out to the reader that the given precision and recall values for the DL approach were calculated on a single-point level. In contrast, the performance measures for the RF classifier given in section 2 are per-tree scores. Figure 4 exemplarily shows the predicted class labels compared to the reference data (”ground truth”) for an area with a size of 6.4 ha. Like the standard method, PointNet++ performed better for coniferous trees than for deciduous trees. One reason for this is the superior representation of the crown shape of coniferous trees in leaf-on condition at a point density of 54 points/m^2. For the time being we just used the geometrical information of the ALS point cloud, whereas the laser intensity has not been included yet. We encountered misclassification effects at the edges of the blocks, since no inter-block neighborhood information was provided to the model. One promising approach to solve this issue is the utilization of Superpoint Graphs (Landrieu et al., 2018) for the semantic labeling of the point cloud. Of course, we expected that the neural network outperforms the RF classifier. Very likely, the errors in the training data and the mentioned edge effects limited a better performance.

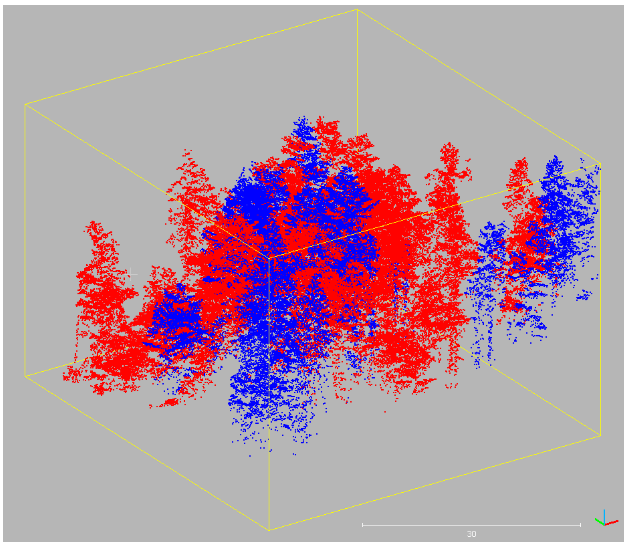

Figure 3. Cubic blocks used for training of PointNet++; 3D points colored in dependence of the class labels ”coniferous” (blue) and ”deciduous” (red)

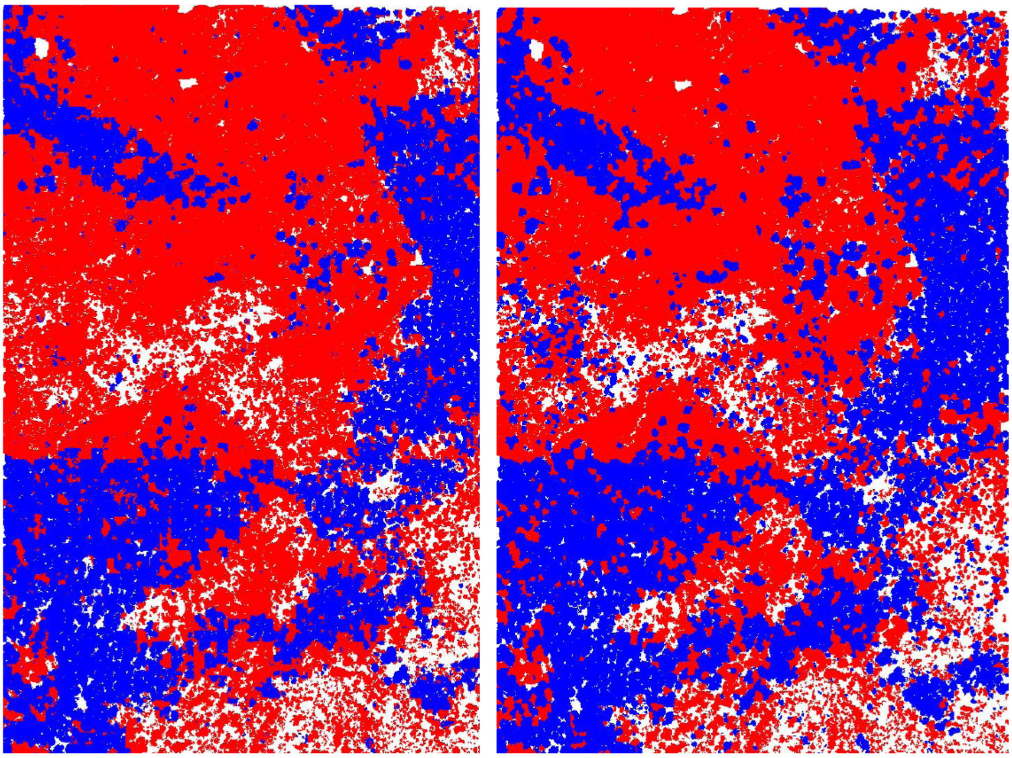

Figure 4. Semantic labeling result (*left*) and reference data (*right*) for coniferous trees (blue) and deciduous trees (red); area size 200 m x 320 m

5. CONCLUSION AND OUTLOOK

The conducted experiments prove that the architecture of the 3D DNN PointNet++ can successfully be adapted to the semantic labeling of large ALS point clouds to map the two tree species spruce and beech. The neural network was trained using ALS data with a point density of 54 points/m^2. The training dataset was generated automatically using a classical feature-based RF classifier that distinguished single coniferous trees (precision = 93%, recall = 80%) and deciduous trees (precision = 82%, recall = 92%). Using the architecture of PointNet++, the achieved classification results for single points belonging to either coniferous trees (precision = 90%, recall = 79%) or deciduous trees (precision = 81%, recall = 91%) are fairly high considering that only the geometry was included. We want to emphasize that - depending on the point density and the extent of the objects - the decisive hyperparameters of the neural network need to be adjusted in order to get a well-performing network for the particular classification task. Moreover, the classification of individual tree species has not been solved yet and it is still a challenging task. For instance, a recent study by Amiri et al. (2019) reported on an accuracy of around 78% by fusing multispectral data from LiDAR and optical imagery to classify four tree species (spruce, fir, beech, dead tree). Hence, future work will focus on the mapping of individual tree species using PointNet++ or comparable DNNs (e.g. SPG, PointCNN). Furthermore, laser intensity values, surface normals and multispectral features should be integrated to further improve the accuracy of the proposed DL approach.

ACKNOWLEDGEMENTS

We would like to thank PD Dr. Marco Heurich and his team from the Bavarian Forest National Park for the supply of the ALS data.

References

Amiri, N., Krzystek, P., Heurich, M., Skidmore, A. K., 2019. Tree species classification by fusing multispectral lidar and aerial imagery. ISPRS Journal of Photogrammetry and Remote Sensing. In review.

Armeni, I., Sener, O., Zamir, A. R., Jiang, H., Brilakis, I., Fischer, M., Savarese, S., 2016. 3d semantic parsing of large-scale indoor spaces. *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition*.

Chang, A. X., Funkhouser, T., Guibas, L., Hanrahan, P., Huang, Q., Li, Z., Savarese, S., Savva, M., Song, S., Su, H. et al., 2015. Shapenet: An information-rich 3d model repository. *arXiv preprint arXiv:1512.03012*.

Dai, A., Chang, A. X., Savva, M., Halber, M., Funkhouser, T., Nießner, M., 2017. Scannet: Richly-annotated 3d reconstructions of indoor scenes. *Proc. Computer Vision and Pattern Recognition (CVPR), IEEE*.

Deng, J., Dong, W., Socher, R., Li, L., Li, K., Fei-Fei, L., 2009. Imagenet: A large-scale hierarchical image database.

Fassnacht, F. E., Latifi, H., Stereńczak, K., Modzelewska, A., Lefsky, M., Waser, L. T., Straub, C., Ghosh, A., 2016. Review of studies on tree species classification from remotely sensed data. *Remote Sensing of Environment*, 186, 64–87.

Geiger, A., Lenz, P., Stiller, C., Urtasun, R., 2013. Vision meets Robotics: The KITTI Dataset. *International Journal of Robotics Research (IJRR)*.

Gevaert, C. M., Persello, C., Nex, F., Vosselman, G., 2018. A deep learning approach to DTM extraction from imagery using rule-based training labels. *ISPRS journal of photogrammetry and remote sensing*, 142, 106–123.

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

Hamraz, H., Jacobs, N. B., Contreras, M. A., Clark, C. H., 2018. Deep learning for conifer/deciduous classification of airborne LiDAR 3D point clouds representing individual trees. *arXiv preprint arXiv:1802.08872*.

He, K., Gkioxari, G., Dollár, P., Girshick, R., 2017. Mask r-cnn. *Proceedings of the IEEE international conference on computer vision*, 2961–2969.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770–778.

Ioannidou, A., Chatzilari, E., Nikolopoulos, S., Kompatsiaris, I., 2017. Deep learning advances in computer vision with 3D data: A survey. *ACM Computing Surveys*, 50.

ISPRS, 2019. Isprs 3d semantic labeling contest. `http://www2.isprs.org/commissions/comm2/wg4/vaihingen-3d-semantic-labeling.html`. Accessed: 2019-03-11.

Krizhevsky, A., Hinton, G. et al., 2009. Learning multiple layers of features from tiny images. Technical report, Citeseer.

Landrieu, L., Simonovsky, M. et al., 2018. Large-scale point cloud semantic segmentation with superpoint graphs. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 4558–4567.

LeCun, Y., Bottou, L., Bengio, Y., Haffner, P. et al., 1998. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86, 2278–2324.

Li, Y., Bu, R., Sun, M., Wu, W., Di, X., Chen, B., 2018. Pointcnn: Convolution on x-transformed points. *Advances in Neural Information Processing Systems*, 828–838.

NVIDIA Corporation, 2019. Nvidia titan v - nvidia's supercomputing gpu architecture. `https://www.nvidia.com/en-us/titan/titan-v/`. Accessed: 2019-03-11.

Qi, C. R., Su, H., Mo, K., Guibas, L. J., 2016. PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation. *arXiv preprint arXiv:1612.00593*.

Qi, C. R., Yi, L., Su, H., Guibas, L. J., 2017. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. *Advances in Neural Information Processing Systems*, 5099–5108.

Qi, C. R., Yi, L., Su, H., Guibas, L. J., 2019. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. `https://github.com/charlesq34/pointnet2`. Accessed: 2019-03-11.

Reitberger, J., Schnörr, C., Krzystek, P., Stilla, U., 2009. 3D segmentation of single trees exploiting full waveform LIDAR data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 64, 561–574.

Simonyan, K., Zisserman, A. et al., 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

Tchapmi, L., Choy, C., Armeni, I., Gwak, J., Savarese, S., 2017. Segcloud: Semantic segmentation of 3d point clouds. *2017 International Conference on 3D Vision (3DV)*, IEEE, 537–547.

Vetrivel, A., Gerke, M., Kerle, N., Nex, F., Vosselman, G., 2018. Disaster damage detection through synergistic use of deep learning and 3D point cloud features derived from very high resolution oblique aerial images, and multiple-kernel-learning. *ISPRS journal of photogrammetry and remote sensing*, 140, 45–59.

Wu, Z., Song, S., Khosla, A., Yu, F., Zhang, L., Tang, X., Xiao, J., 2015. 3d shapenets: A deep representation for volumetric shapes. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 1912–1920.

Zhao, R., Pang, M., Wang, J., 2018. Classifying airborne LiDAR point clouds via deep features learned by a multi-scale convolutional neural network. *International Journal of Geographical Information Science*, 32, 960–979.

Zhou, Y., Tuzel, O. et al., 2018. Voxelnet: End-to-end learning for point cloud based 3d object detection. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 4490–4499.