Transfer learning for indoor object classification

From images to point clouds

Balado, J.; Díaz-Vilariño, L.; Verbree, E.; Arias, P.

DOI
10.5194/isprs-Annals-V-4-2020-65-2020

Publication date
2020

Document Version
Final published version

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

TRANSFER LEARNING FOR INDOOR OBJECT CLASSIFICATION: FROM IMAGES TO POINT CLOUDS

J. Balado[a,b*], L. Díaz-Vilariño[a,b], E.Verbree[b], P. Arias[a]

[a]Universidade de Vigo. CINTECX, Applied Geotechnologies Research Group.
Campus universitario de Vigo, As Lagoas, Marcosende 36310 Vigo, Spain
(jbalado, lucia, parias)@uvigo.es

[b]Delft University of Technology, Faculty of Architecture and the Built Environment, GIS Technology Section.
2628 BL Delft, The Netherlands
(J.BaladoFrias, L.Diaz-Vilarino, E.Verbree)@ tudelft.nl

Commission IV/5

KEY WORDS: Deep Learning, data augmentation, Convolutional Neural Networks, indoor environments, InceptionV3, multi-view.

ABSTRACT:

Indoor furniture is of great relevance to building occupants in everyday life. Furniture occupies space in the building, gives comfort, establishes order in rooms and locates services and activities. Furniture is not always static; the rooms can be reorganized according to the needs. Keeping the building models up to date with the current furniture is key to work with indoor environments. Laser scanning technology can acquire indoor environments in a fast and precise way, and recent artificial intelligence techniques can classify correctly the objects that contain. The objective of this work is to study how to minimize the use of point cloud samples in Neural Network training, tedious to label, and replace them with images obtained from online sources. For this, point clouds are converted to images by means of rotations and projections. The conversion of a 3D vector data to a 2D raster allows the use of Convolutional Neural Networks, the achievement of several images for each acquired point cloud object and the combination with images obtained from online sources, such as Google Images. The images have been distributed among the validation and testing training sets following different percentages. The results show that, although point cloud images cannot be completely dispensed within the training set, only 10% of these achieve high accuracy in the classification.

1. INTRODUCTION

Furniture is a key element of indoor environments. These objects allow people and autonomous robots to interact with buildings, locate services and tools, and recognize spaces based on the type of objects they contain. Some models, such as the CityGML standard at its highest level of detail (Biljecki et al., 2016), integrate objects within buildings to know the space occupied and services available. Indoor environments are also changing, rooms are usually reorganized and adapted to current needs. Therefore, it is essential to provide methods to acquire and map these objects quickly and minimize manual intervention.

Indoor laser scanning technology has evolved significantly in recent years. The platforms where the laser scanner is mounted have been diversified into trolleys (Chen et al., 2019), backpacks (Rönnholm et al., 2015), manual tools (Maboudi et al., 2017), mixed reality devices (Khoshelham et al., 2019), robots (Frías et al., 2019), etc. These ramifications allow indoor environments can be acquired more quickly than with conventional Terrestrial Laser Scanning, thus obtaining more data. However, this data is often not enough and must be labelled if Deep Learning (DL) technologies are implemented. Therefore, the task of acquiring and labeling samples is a time-consuming manual process. Although there are datasets with indoor labelled point clouds (Uy et al., 2019), this data does not always match the user's needs, or the number of samples is low to employ certain techniques.

The objective of this work is to evaluate the use of images of indoor objects to minimize the number of point clouds needed in the training of Convolutional Neural Networks (CNN). Images are easier and faster to obtain and label compared to point clouds, and the objects maintain a clear relation in both images and clouds. Different training sessions are held where the percentage varies between images obtained from online sources and images obtained from point clouds.

The rest of this paper is organized as follows. Section 2 collects related work about object classification with Machine Learning (ML) techniques. Section 3 presents an overview of the designed method. Section 4 is devoted to analyse the results. Finally, Section 5 concludes this work.

2. RELATED WORK

Object classification is a well-studied topic, both in point clouds and in images. Many of the object classification techniques can be applied indoors and outdoors indistinctly (Balado et al., 2020). Objects in point clouds can be classified with ML techniques by feature extraction, converting point clouds into 2D or 3D images or using point cloud-based neural networks.

* Corresponding author

ML techniques need a low number of samples for training, compared with DL techniques. ML techniques must be designed to extract the most relevant point cloud object features. The choice of features is a design decision, so depending on the design knowledge, relevant features can be lost and other features less relevant can be added. A tendency in the use of these techniques is to extract all available features and let the classifier detect those that are relevant. ML classifiers, such as SVM, Random Forest, Trees, etc., obtain good results in non-complex problems, with low computational cost and little time in dataset generation. Lai and Fox, (2010) extract features from Google's 3D Warehouse to obtain more data samples. Roynard et al., (2016) uses 991 features to train a Random Forest classifier. Oesau et al., (2016) transform point clouds objects to histograms via planar abstraction.

Object classification in images with 2D-CNN is one of the most widespread research lines today. There is a wide variety of network architectures available, implementation is quick and does not require a deep understanding of the problem to be addressed. When generating 2D samples from 3D data, data augmentation with object rotations can be implemented, thus significantly minimizing the number of acquired objects required for training (Tchapmi et al., 2017). The main drawback is that the 3D to 2D conversion loses one data dimension. To minimize this, some authors choose to use orthogonal sections of the object (Gomez-Donoso et al., 2017) and others transform the cloud into depth images (Pang and Neumann, 2016).

The first network to address the problem of classification directly in 3D was VoxNet (Maturana and Scherer, 2015). This network uses 32x32x32 voxels as input, so the point cloud must be structured into a 3D image. The main problem when adapting vector data to 32 levels in each dimension is the resolution loss and the empty voxel generation. In addition, some authors consider that 2D-CNN with multi-views obtain better results than these 3D-CNN (Griffiths and Boehm, 2019; Qi et al., 2016b).

Recently, some authors have designed network architectures that use point clouds as input. These architectures are based on spatial relationships (Qi et al., 2017, 2016a) and graph theory (Feng et al., 2019; Wang et al., 2018). The strong point of these networks is no information is lost due to point cloud conversion to other formats. Their weak point is that they need a much higher computational cost than the alternatives. Garcia-Garcia et al., (2016) train PointNet with CAD models of objects to classify them. Wu et al., (2019) employ synthetic data from videogames to train SqueezeSegV2.

	ML e.g. SVM	2D-CNN e.g. ResNet	3D-CNN e.g. VoxNet	Points e.g. PointNet
Problem abstraction	Low	High	High	High
Number of samples	Low	High	High	High
Data augmentation*	No	Yes	Yes	No
Computational cost	Low	Medium	Medium-High	High

Table 1. Comparison between different Artificial Intelligence object classification methods for point clouds (*Data augmentation refers to generating several samples per object by rotations, not by adding noise)

With regard to the mentioned works, briefly compared in Table 1, the method presented in this paper opts for the conversion of point clouds to images in order to use a 2D-CNN network. The decision is substantiated in the following reasons: (1) The shape of the object is preserved, one of the most relevant factors at classification. (2) It allows the use of data augmentation, generating multiple samples per object. (3) Computation time and cost are reduced compared to 3D techniques. (4) Existing 2D networks are better optimized than their 3D equivalents and manual feature extraction techniques. (5) Point cloud images can be combined with images obtained from online sources.

3. METHOD

The classification is based on images downloaded from online sources and images generated from point clouds (hereinafter called point cloud images). Depending on the number of samples per class, multi-view data augmentation is applied to obtain enough samples to evaluate training and assess the behavior of the algorithm. The samples are then distributed among the training, validation, and testing sets (Figure 1). In this section, the generation of images from point clouds, the CNN selection and the adaptation of the images are explained.

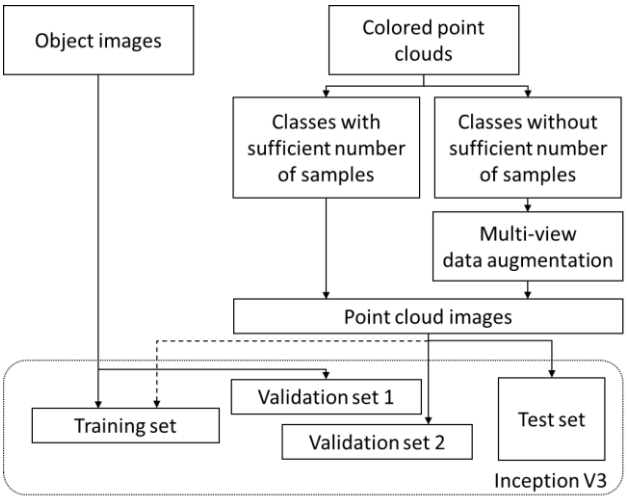


Figure 1. Workflow

3.1 Image generation from point clouds

The input data are individualized point clouds of objects $P = [X\ Y\ Z\ R\ G\ B]$, where the first three columns are 3D coordinate and the last three are color information. The conversion from point clouds to images is done through an isometric projection. The point cloud is distributed in a plane, which can be visualized and saved as an image. In pixels where more than one point is projected, the color assigned is the average color of the corresponding points. White color is assigned to pixels without points. A point cloud rasterization (Balado et al., 2017) is not necessary since an aspect ratio is not maintained when adapting images to the CNN entrance.

If it is necessary to rotate the point cloud *P* to generate multiple views of the same point cloud object (data augmentation), a rotation is executed with an angle resolution *r* on Z axis. Equation 1 shows the rotation matrix on the Z axis according to a step *i* of angle *r*. The number of rotations coincides with the number of final images per object. In this way, multiple images can be

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-V-4-2020-65-2020 | © Authors 2020. CC BY 4.0 License.

created per object, as long as the angle r ensures that images of the same object are sufficiently distinct.

$$P_{Ri} = PR_i = [X\ Y\ Z]\begin{bmatrix}\cos(i*r) & -\sin(i*r) & 0\\ \sin(i*r) & \cos(i*r) & 0\\ 0 & 0 & 1\end{bmatrix} \quad (1)$$

For the visualization of the object in isometric projection, and after rotating the object if multi-view generation is necessary, a rotation of 30 degrees is executed on the axis Y according to Equation 2. Then, point cloud is projected on the X plane a by removing the attribute X.

$$P_{Pi} = P_{Ri}R_P = P_{Ri}\begin{bmatrix}1 & 0 & 0\\ 0 & \cos(30º) & -\sin(30º)\\ 0 & \sin(30º) & \cos(30º)\end{bmatrix} \quad (2)$$

3.2 Classification

The InceptionV3 architecture (Szegedy et al., 2016) is used for the classification as it is one of the networks with the best accuracy in relation to the rate of operations required for their training (Canziani et al., 2016). This architecture has proven to work well in a multitude of object classification applications (Saini and Susan, 2019; Xia et al., 2017). The InceptionV3 network has an input size of 299x299x3 pixels. Since the images obtained from online sources and the images obtained from point clouds are in RGB color format, there is no need to adjust color channels. Since all images have different sizes, the images are resized to fit the network input (Gao and Gruev, 2011). Color assignment is performed by bicubic interpolation; the output pixel value is a weighted average of pixels in the four vicinity.

4. RESULTS AND DISCUSSION

4.1 Data

The point clouds used for training, validation, and testing of the neural network were obtained from areas 1 to 4 of the 2D-3D Stanford Dataset (Armeni et al., 2017). The dataset contains indoor point clouds colored in RGB. The classes of furniture and number of objects available in the dataset are 56 boards, 179 bookshelves, 676 chairs, 21 sofas and 145 tables. Objects have an average density of 10 thousand points per square meter. The number of samples among classes is clearly unbalanced. For each class, 200 point cloud images were generated following the abovementioned method (projection and data augmentation). For each class, 550 images were downloaded from Google Images using the "Download All Images" extension. Figure 2 shows samples for each class.

4.2 Training

Once sufficient samples for each class were available, they were distributed and CNN was trained. For each class, 500 samples were used for training, 50 for validation and 100 for testing. The training set consists of 500 images, of which a small percentage (between 0 to 10%) corresponds to point cloud images, the complementary images are downloaded images (respectively 100% to 90%). This variation was done in 2% increments (10 samples per class). Given the limited and unbalanced number of point cloud objects, it was not possible to create a training set with only point cloud images. There were two different validation sets, one consists of 50 samples obtained from downloaded images and other consists of 50 samples from point cloud images. With the different combinations between training and validation sets, a total of 12 training sessions have been carried out, 6 with each validation set. The test dataset was 500 point cloud images (100 samples per class).

The hyperparameters of the training were: optimization method sgdm, learning rate 0.0001, Momentum 0.9, L2 Regularization 0.0001, Max Epochs 10 and Mini Batch Size 16. Each training session took approximately 55 minutes. The method was implemented in Matlab and processed on an Intel Core i7-7700HQ CPU 2.80 GHz with 16 GB RAM.

Figure 3 and Figure 4 show the evolution of the loss in the successive training sessions containing online images and point cloud images in the validation set respectively. All the networks have converged satisfactorily, however, those that use online images as validation set shows a faster convergence since they do not consider the same feature selection of point cloud images.

4.3 Results and discussion

Table 2 and Table 3 compile the results obtained from the different training sessions on the testing set. Figure 5 shows images of correctly classified objects. Without any point cloud

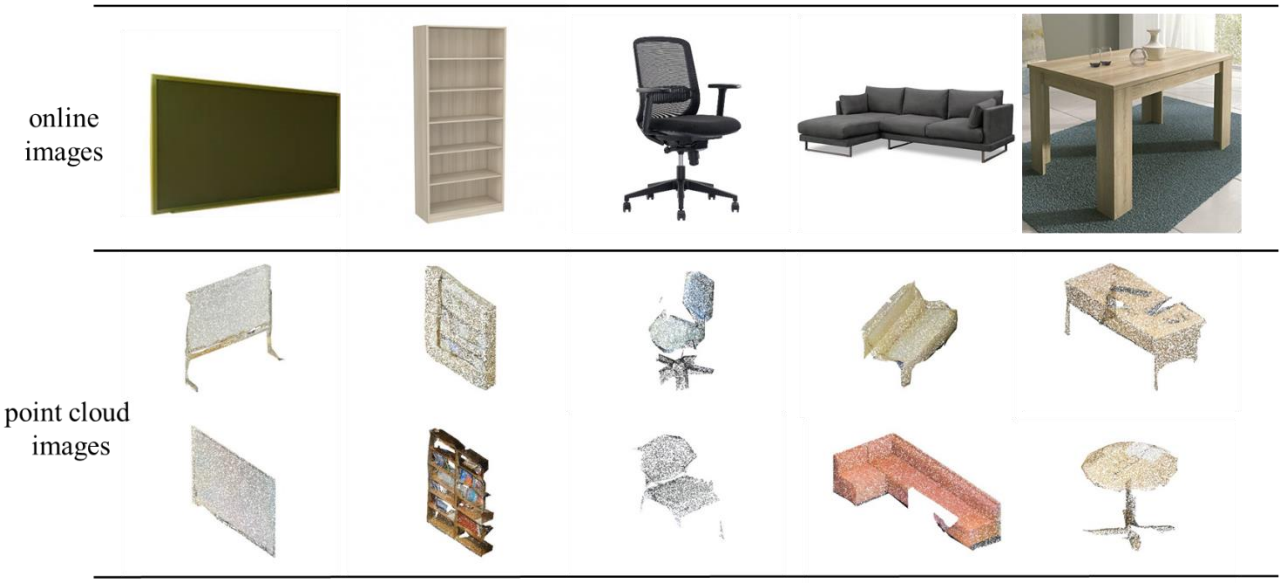


Figure 2. Samples of the five classes: above, images obtained from online sources; below, images obtained from point clouds.

image in the training set (0% of point cloud samples), the neural network was unable to learn appropriate features to identify each object. Therefore, point cloud images colored in RGB were not similar enough to online images to obtain a satisfactory classification. Adding point cloud images in the training set improves the accuracy. The first ingestion of 10 samples per object (2% of point cloud samples) in the training set increased the accuracy by twofold to 0.67. As point cloud images continued to be introduced into the training set, accuracy increased steeply to 0.88 and 0.87, depending on the validation set. with 50 samples per object (10% of point cloud samples in the training set). This accuracy positions the proposed method with the minimization of point cloud objects very close to the state of the art in 2D-3D Stanford Dataset (Turkoglu et al., 2018), and even improving others (McCormac et al., 2017; Tchapmi et al., 2017; Turkoglu et al., 2018). However, these works present semantic segmentation methods of indoor environment point clouds, and not only object classification method as proposed here, that would require a previous phase of object segmentation from structural elements and their individualization.

Between the use or not of point cloud images in the validation set, no great accuracy differences have been observed. Point cloud images can be eliminated from the validation set to reduce the number of point cloud samples.

Table 4 and Table 5 show the confusion matrices for training sessions with 10% of point cloud images. The classes with the highest accuracy were board and chair. From the analysis of the images and errors, the causes of the most relevant confusions can be deduced. Bookshelves were confused with other objects because of their great variation in forms, textures, and contents. Sofas had a high confusion with chairs since in the set of chairs there are some easy chairs. Finally, tables include tables of different shapes as well as desks; in most cases, tables have objects on top of them that difficult visualization. It has also been observed that the objects in point cloud images contained some errors caused in the acquisition and subsequent representation that may influence training and classification. These point clouds often presented diffuse contours, differences in density between objects and between areas of the same object and strong occlusions (Figure 6). Noise can create shapes that confuse CNN. Occlusions can hide object shapes that the CNN needs for object identification.

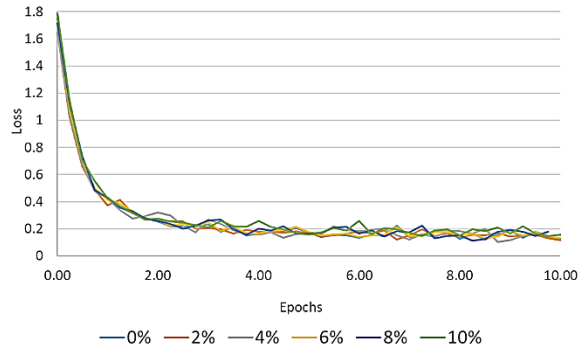


Figure 3. Loss evolution with different percentage of point cloud images in training set and online images in validation set.

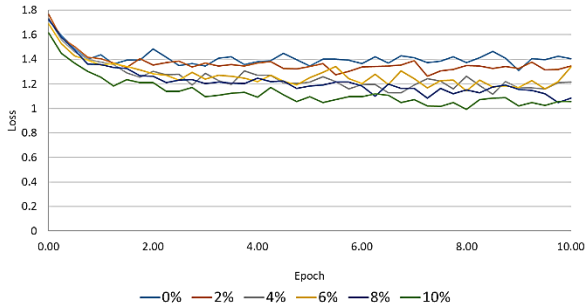


Figure 4. Loss evolution with different percentage of point cloud images in training set and point cloud images in validation set.

	0%	2%	4%	6%	8%	10%
board	0.53	0.88	0.92	0.89	0.98	0.96
shelves	0.51	0.61	0.69	0.63	0.75	0.75
chair	0.19	0.69	0.93	0.96	0.98	0.99
sofa	0.26	0.71	0.59	0.83	0.76	0.88
table	0.19	0.49	0.57	0.68	0.77	0.83
TOTAL	**0.34**	**0.68**	**0.74**	**0.80**	**0.85**	**0.88**

Table 2. Evaluation of accuracy by class according to the percentage of point cloud images in the training set with online images in the validation set.

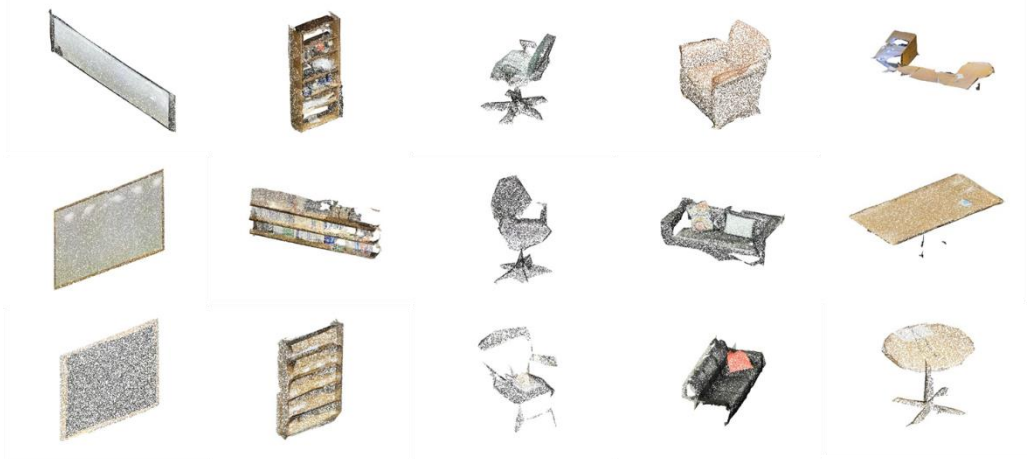


Figure 5. Samples correctly classified by the trained InceptionV3 with 10% of point cloud samples in training set and with online images in the validation set.

	0%	2%	4%	6%	8%	10%
board	0.52	0.90	0.94	0.90	0.97	0.98
shelves	0.55	0.73	0.74	0.70	0.68	0.73
chair	0.15	0.67	0.72	0.93	0.96	0.99
sofa	0.41	0.66	0.87	0.92	0.82	0.88
table	0.20	0.40	0.62	0.71	0.73	0.77
TOTAL	**0.37**	**0.67**	**0.78**	**0.83**	**0.83**	**0.87**

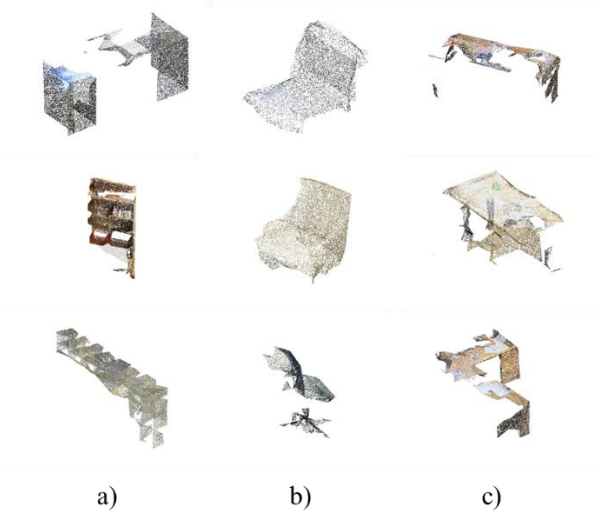
Table 3. Evaluation of accuracy by class according to the percentage of point cloud images in the training set with point cloud images in the validation set.

ref\pred	board	shelves	chair	sofa	table
board	**96**	1	0	0	3
shelves	4	**75**	6	4	11
chair	0	0	**99**	0	1
sofa	0	1	10	**88**	1
table	2	4	7	4	**83**

Table 4. Confusion matrix of CNN trained with online images in the validation set.

ref\pred	board	shelves	chair	sofa	table
board	**98**	0	0	0	2
shelves	7	**73**	9	5	6
chair	0	0	**99**	1	0
sofa	0	4	7	**88**	1
table	4	4	10	5	**77**

Table 5. Confusion matrix of CNN trained with point cloud images in the validation set.



a) b) c)

Figure 6. Samples with strong changes in intensity, occlusions and shape variation: a) bookshelves, b) chairs and c) tables

5. CONCLUSIONS

In this work, the use of online images has been studied to minimize the number of point cloud samples needed to train a neural network to the classification of indoor objects. Classification with a CNN has been adopted, so point clouds have been converted into images. Several training sets have been designed where the percentage of samples obtained from point clouds and online images is varied.

Colored point clouds provided by the 2D-3D Stanford Dataset and images from online sources were used to classify five classes of indoor objects. The results show that online images cannot be used exclusively to train a CNN whose objective is to classify point clouds (even if these have color). The accuracy of the classifier increases gradually as the number of images obtained from point clouds in the training set increases. With 10% of point cloud images in the training set, an accuracy of 0.88 was achieved. Although the proposed method minimizes the number of point cloud samples, the choice of how many samples to use in the training is at the disposal of the creator of the dataset, the number of available samples and the final accuracy desired. Future work will focus on studying how occlusions and other anomalies in point clouds of objects affect classification results.

ACKNOWLEDGEMENTS

Authors would like to thank to the Xunta de Galicia given through human resources grant (ED481B-2019-061, ED481D 2019/020) and competitive reference groups (ED431C 2016-038), the Ministerio de Ciencia, Innovación y Universidades - Gobierno de España- (RTI2018-095893-B-C21). 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 views of the author(s). Neither the Innovation and Networks Executive Agency (INEA) or the European Commission is in any way responsible for any use that may be made of the information it contains. The statements made herein are solely the responsibility of the authors.

Travel expenses were partially covered by the Travel Award sponsored by the open access journal ISPRS International Journal of Geo-Information published by MDPI.

REFERENCES

Armeni, I., Sax, A., Zamir, A. ~R., Savarese, S., 2017. Joint 2D-3D-Semantic Data for Indoor Scene Understanding. ArXiv e-prints.

Balado, J., Díaz-Vilariño, L., Arias, P., Garrido, I., 2017. Point Clouds To Indoor / Outdoor Accessibility Diagnosis. ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci. ISPRS Geospatial Week 2017 IV-2/W4, 18–22. https://doi.org/doi.org/10.5194/isprs-annals-IV-2-W4-287-2017

Balado, J., Sousa, R., Díaz-Vilariño, L., Arias, P., 2020. Transfer Learning in urban object classification: Online images to recognize point clouds. Autom. Constr. 111, 103058. https://doi.org/https://doi.org/10.1016/j.autcon.2019.103058

Biljecki, F., Ledoux, H., Stoter, J., 2016. An improved LOD specification for 3D building models. Comput. Environ. Urban Syst. 59, 25–37. https://doi.org/https://doi.org/10.1016/j.compenvurbsys.2016.04.005

Canziani, A., Paszke, A., Culurciello, E., 2016. An Analysis of Deep Neural Network Models for Practical Applications.

Chen, C., Tang, L., Hancock, C., Zhang, P., 2019. Development of low-cost mobile laser scanning for 3D construction indoor mapping by using inertial measurement unit, ultra-wide band and 2D laser scanner. Eng. Constr. Archit. Manag. 26, 1367–1386.

Feng, Y., You, H., Zhang, Z., Ji, R., Gao, Y., 2019. Hypergraph Neural Networks, in: AAAI Technical Track: Machine Learning.

https://doi.org/https://doi.org/10.1609/aaai.v33i01.33013558

Frías, E., Díaz-Vilariño, L., Balado, J., Lorenzo, H., 2019. From BIM to Scan Planning and Optimization for Construction Control. Remote Sens. . https://doi.org/10.3390/rs11171963

Gao, S., Gruev, V., 2011. Bilinear and bicubic interpolation methods for division of focal plane polarimeters. Opt. Express 19, 26161–26173. https://doi.org/10.1364/OE.19.026161

Garcia-Garcia, A., Gomez-Donoso, F., Garcia-Rodriguez, J., Orts-Escolano, S., Cazorla, M., Azorin-Lopez, J., 2016. PointNet: A 3D Convolutional Neural Network for real-time object class recognition, in: 2016 International Joint Conference on Neural Networks (IJCNN). pp. 1578–1584. https://doi.org/10.1109/IJCNN.2016.7727386

Gomez-Donoso, F., Garcia-Garcia, A., Garcia-Rodriguez, J., Orts-Escolano, S., Cazorla, M., 2017. LonchaNet: A sliced-based CNN architecture for real-time 3D object recognition, in: 2017 International Joint Conference on Neural Networks (IJCNN). pp. 412–418. https://doi.org/10.1109/IJCNN.2017.7965883

Griffiths, D., Boehm, J., 2019. A Review on Deep Learning Techniques for 3D Sensed Data Classification. Remote Sens. . https://doi.org/10.3390/rs11121499

Khoshelham, K., Tran, H., Acharya, D., 2019. INDOOR MAPPING EYEWEAR: GEOMETRIC EVALUATION OF SPATIAL MAPPING CAPABILITY OF HOLOLENS. ISPRS - Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. XLII-2/W13, 805–810. https://doi.org/10.5194/isprs-archives-XLII-2-W13-805-2019

Lai, K., Fox, D., 2010. Object Recognition in 3D Point Clouds Using Web Data and Domain Adaptation. Int. J. Rob. Res. 29, 1019–1037. https://doi.org/10.1177/0278364910369190

Maboudi, M., Bánhidi, D., Gerke, M., 2017. Evaluation of indoor mobile mapping systems.

Maturana, D., Scherer, S., 2015. VoxNet: A 3D Convolutional Neural Network for real-time object recognition, in: 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). pp. 922–928. https://doi.org/10.1109/IROS.2015.7353481

McCormac, J., Handa, A., Leutenegger, S., Davison, A.J., 2017. SceneNet RGB-D: Can 5M Synthetic Images Beat Generic ImageNet Pre-training on Indoor Segmentation?, in: 2017 IEEE International Conference on Computer Vision (ICCV). pp. 2697–2706. https://doi.org/10.1109/ICCV.2017.292

Oesau, S., Lafarge, F., Alliez, P., 2016. Object classification via planar abstraction, in: ISPRS Congress. Prague, Czech Republic.

Pang, G., Neumann, U., 2016. 3D point cloud object detection with multi-view convolutional neural network, in: 2016 23rd International Conference on Pattern Recognition (ICPR). pp. 585–590. https://doi.org/10.1109/ICPR.2016.7899697

Qi, C.R., Su, H., Mo, K., Guibas, L.J., 2016a. PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation. CoRR abs/1612.0.

Qi, C.R., Su, H., Nießner, M., Dai, A., Yan, M., Guibas, L.J., 2016b. Volumetric and Multi-View CNNs for Object Classification on 3D Data. IEEE Conf. Comput. Vis. Pattern Recognit. abs/1604.0, 5648–5656. https://doi.org/10.1109/CVPR.2016.609

Qi, C.R., Yi, L., Su, H., Guibas, L.J., 2017. PointNet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space. CoRR abs/1706.0.

Rönnholm, P., Kukko, A., Liang, X., Hyyppä, J., 2015. Filtering the outliers from backpack mobile laser scanning data. Photogramm. J. Finl. 24, 20–34. https://doi.org/10.17690/015242.2

Roynard, X., Deschaud, J.-E., Goulette, F., 2016. Fast and Robust Segmentation and Classification for Change Detection in Urban Point Clouds. ISPRS - Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. XLI-B3, 693–699. https://doi.org/10.5194/isprsarchives-XLI-B3-693-2016

Saini, M., Susan, S., 2019. Data Augmentation of Minority Class with Transfer Learning for Classification of Imbalanced Breast Cancer Dataset Using Inception-V3 BT - Pattern Recognition and Image Analysis, in: Morales, A., Fierrez, J., Sánchez, J.S., Ribeiro, B. (Eds.), . Springer International Publishing, Cham, pp. 409–420.

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z.B., 2016. Rethinking the Inception Architecture for Computer Vision, in: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 2818–2826. https://doi.org/10.1109/CVPR.2016.308

Tchapmi, L., Choy, C., Armeni, I., Gwak, J., Savarese, S., 2017. SEGCloud: Semantic Segmentation of 3D Point Clouds, in: 2017 International Conference on 3D Vision (3DV). pp. 537–547. https://doi.org/10.1109/3DV.2017.00067

Turkoglu, M.O., Haar, F.B. Ter, Stap, N. van der, 2018. Incremental Learning-Based Adaptive Object Recognition for Mobile Robots, in: 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). pp. 6263–6268. https://doi.org/10.1109/IROS.2018.8593810

Uy, M.A., Pham, Q.-H., Hua, B.-S., Nguyen, T., Yeung, S.-K., 2019. Revisiting Point Cloud Classification: A New Benchmark Dataset and Classification Model on Real-World Data, in: The IEEE International Conference on Computer Vision (ICCV).

Wang, Y., Sun, Y., Liu, Z., Sarma, S.E., Bronstein, M.M., Solomon, J.M., 2018. Dynamic Graph {CNN} for Learning on Point Clouds. ACM Trans. Graph. 38, 146:1--146:12. https://doi.org/10.1145/3326362

Wu, B., Zhou, X., Zhao, S., Yue, X., Keutzer, K., 2019. SqueezeSegV2: Improved Model Structure and Unsupervised Domain Adaptation for Road-Object Segmentation from a LiDAR Point Cloud, in: 2019 International Conference on Robotics and Automation (ICRA). pp. 4376–4382. https://doi.org/10.1109/ICRA.2019.8793495

Xia, X., Xu, C., Nan, B., 2017. Inception-v3 for flower classification, in: 2017 2nd International Conference on Image, Vision and Computing (ICIVC). pp. 783–787. https://doi.org/10.1109/ICIVC.2017.7984661