Skoltech
Skolkovo Institute of Science and Technology

Thesis Changes Log

Name of Candidate: Emre Özdemir

PhD Program: Engineering Systems

Title of Thesis: Point Cloud Classification for Geospatial Applications

Supervisor: Prof. Alessandro Golkar, Dr. Fabio Remondino

The thesis document includes the following changes in answer to the external review process.

I have corrected numerous typos and misprints as well as refined the bibliography list, as was pointed out by Jury members. Please find the changes log below:

All Pages
Sentences with we/us/our revised and these words removed.

Cover
Page 1
- "GEOSPATIAL POINT CLOUD CLASSIFICATION" ---> "A DEEP LEARNING FRAMEWORK FOR GEOSPATIAL POINT CLOUD CLASSIFICATION"
- "Professor Alessandro Golkar" ---> "Associate Professor Alessandro Golkar"

Abstract
Page 6
- "(ii) Generalization of the learned classification ability (i.e., predicting on datasets with acquisition setup)" ---> "(ii) A high generalization ability (i.e., predicting on datasets with different acquisition setup)"
- **"…** with the current …**"** ---> "… compared to the current"

List of Symbols, Abbreviations
Page 17
- Added: "DEM – Digital Elevation Model"
- Removed: "DTM – Digital Terrain Model"

Page 18
- Added: "TIN – Triangulated Irregular Network"

Chapter 1
Page 33
- "Objects on maps…" ---> "Objects in maps…"

Page 35
- "…to either specific data (e.g., only for LiDAR acquisitions) or …" ---> "…either to specific data (e.g., only for LiDAR acquisitions) or to specific …"
- "…issues to be faced in the PhD…" ---> "…problems to be faced in this PhD…"

Page 36
- "… open issues." ---> "… open problems."

Page 37
- "…land surveying study, …" ---> "…land survey, …"

Page 41

- "…being a grid-based height representation (structured …" ---> "…being a height representation (i.e., structured …"
- "…on it, can be…" ---> "…on it. They can be…"
- "… DTM (blue) …" ---> "… DEM (blue) …"

Page 43

- "In geospatial science, a point cloud is commonly 3D and defined in a projected coordinate system such as UTM." ---> "In geospatial science, a 3D point cloud is defined in a projected -Cartesian- coordinate system such as UTM."
- "There are two commonly used geospatial point cloud types;…" ---> "There are two common data types representing 3D geospatial data;…"
- "… a grid structure. This means, to produce…" ---> "… as a grid structure, or a surface model (i.e., Triangulated Irregular Network, TIN). For instance, to produce …"
- "...represents the height. In Figure 1-7,..." ---> "...represents the height. This procedure is also known as rasterization. In Figure 1-7,..."
- "… DSM (Imst, …" ---> "a 2.5D DSM grid (Imst, …"

Page 44

- "…(left) and DSM, …" ---> "…(left) and DSM (right), …"
- "Thus, in order to achieve a 3D object space, there should…" ---> "In order to reconstruct an object in 3D space, there must..."

Page 45

- Removed: "Using a single photograph with photogrammetry technique will result in a scaled version of that photograph, which means the result will be in a 2D object space."
- "through *C*. Therefore, *P'*, *C* and P_{xyz} are collinear points, as well as *P"*, *C* and P_{xyz}." --->"…through *C' and C"*. Therefore, *P'*, *C'* and P_{xyz} are collinear points, as well as *P"*, *C"* and P_{xyz}."

Page 46

- Figure 1-9 updated with projection centers C' and C''

Page 47

- "Equation **Error! Reference source not found.**, where r_{ij}…" ---> "Equation **Error! Reference source not found.**, where ξ, η are the two-dimensional image coordinates, r_{ij}…"

Page 48

- "… five images are acquired simultaneously: …" ---> "… five images are acquired: …"

Page 49

- "… not allow image acquisition planning by experts." ---> "… not allow image acquisition planning by experts after the initial system design."
- "… difference is the atmospheric effects." ---> "… difference is in the atmospheric effects."

Page 51

- "Example usage of semantic information and …" ---> "An example for semantic information extraction and …"
- Figure 1-12 "… for each planar segment…" ---> "for each extracted planar segment".

Page 52

- "Classification can be …" ---> "Point-wise classification can be …"
- "… can be referred to with different names." ---> "… can be referred to by different names."

Page 53

- "- 3D shape classification is where the DNN learns the global shape of the given point cloud objects (i.e., a teapot, a car)." ---> "- 3D shape classification is where the deep neural network (DNN) learns the global shape of the given point cloud objects (i.e., a teapot, a car)."
- Added: "*3D point cloud instance segmentation* is similar to the 3D point cloud classification. The main difference is that with instance segmentation, the object instances (i.e., individual buildings) are distinguished as well."

Page 54

- **"…**unclassified (b) point clouds**…" ---> "…**unclassified (b) 3D point clouds**…"**
- "Given the motivations (Section 1.1) and the recent developments in the state-of-the-art (Chapter 2) the goal of this study is to develop a framework that achieves better or similar accuracies compared to the state-of-the-art ($\gtrsim$80%) with a more efficient (in terms of computational time and hardware

requirements) methodology.” ---> “Given the motivations (Section 1.1) and the recent developments in the state-of-the-art (Chapter 2) the goal of this study is to develop a point cloud classification framework for geospatial point clouds that achieves better or similar accuracies compared to the state-of-the-art (OA $\gtrsim$80%) with a more efficient (in terms of computational time and hardware requirements) methodology.”

Page 55

- Added: “The scientific objectives of this thesis will be discussed further at the end of this section, after discussing the objectives, innovation and novelty aspects of the developed framework.”
- “…towards daily applications.” ---> “…towards daily applications (such as city-scale or country-scale applications of a national mapping agency).”

Page 56

- “a 2D matrix or a 3D array per point” ---> “a 2D or a 3D data array per point”
- "...generalization capability." ---> "...generalization capability, which supports the method towards being a feasible solution for daily applications as it would not need separate training dataset for each distinct dataset."

Page 57

- Added: “The main contribution of this thesis is developing a new perspective, in which classical machine learning (i.e., extracting handcrafted features and feeding these to a classifier) and deep learning (i.e., use of deep neural networks that learns features by itself) approaches are combined by transforming 3D point clouds (i.e., irregularly distributed points in 3D cartesian coordinate system) into data arrays (consist of handcrafted features) that can be fed to CNNs, which allows computationally efficient and highly accurate classification of point clouds.
  Therefore, objectives of this thesis can be seen as researching (Chapter 2), proposing and development (Chapter 3) of a framework that actualizes this perspective, as well as evaluating (Chapter 4) and benchmarking (Chapter 5) it.”

Page 59

- “An Overview of the Artificial Intelligence” ---> “An Overview of Artificial Intelligence”.

Chapter 2

Page 65

- “…while this is not the case for ML” ---> “…while this is not the case for classic ML”
- “ML methods for geospatial point cloud classification focus on labeling each point individually by processing their features.” ---> “Classic ML methods for geospatial point cloud classification commonly focus on labeling each point individually by processing their features.”

Page 66

- “Handcrafted features typically describe local (i.e., planarity, linearity) or global geometry (i.e., height above ground level, surface normal).” ---> “Handcrafted features typically describe local geometry (i.e., planarity, linearity, height above ground level, surface normal).”
- Added: “Classical ML methods commonly have less parameters to train, which requires less training data.”

Page 67

- “spherical neighborhoods for indoor and outdoor” ---> “spherical neighborhoods for feature extraction in indoor and outdoor”

Page 68

- “… can achieve high accuracies such as…” ---> “… can achieve accuracies such as…”.

Page 69

- “… different data processing necessities.” ---> “…different applications.”

Page 70

- “… are incapable of handling…” ---> “… are incapable of handling such amount of data…”
- “…training and validation…” ---> “training and evaluation”

Page 74

- Removed: “The candidate objects’ midpoints are then compared with a 30cm threshold in order to merge them to a single object or leave them as separate objects.”

Chapter 3

Page 79

- Figure 3-1 updated by inclusion of 'Point Cloud' for 'Input' and 'Output'
- Added: "The framework design is based on the literature, which proves the classification capabilities of the CNNs, as well as the use of handcrafted features in the classical ML domain. The idea behind the framework design is bringing these two approaches together, which allows utilizing a DNN with few layers and parameters for computational efficiency."
- Added: "The framework receives 3D point clouds and outputs class labels per-point. If these point clouds include any radiometric data (i.e., RGB color) or LiDAR features (i.e., intensity, number of returns) these data are exploited, as well."

Page 81

- "Using the voxel-grid filtering, also the noise is reduced, as output is the voxel centers instead of the points' centroid (red point in **Error! Reference source not found.**). As the output…" ---> "Using the voxel-grid filtering, also the noise is reduced. Besides, the output…"

Page 82

- Figure 3-3 Images updated to match with the legend.

Page 83

- "when there are very few points (i.e., 2 points) within the defined search…" ---> "when there are very few points (i.e., less than three points) within the defined 3D spherical search…"

Page 84

- Algorithm 3-1: "**Input:** 3D Point Cloud" ---> "**Input:** 3D Point Cloud, Number of Scales, Number of Points per Scale"

Page 86

- Figure 3-4 Legend added showing color scale low to high values.
- "Local planarity is the average distance between the neighboring points (p_k) to the best-fit plane ($P^{\rightarrow}$) of these neighborhood points." ---> "Local planarity is the mean distance between the neighboring points (p_k) to the best-fit plane ($P^{\rightarrow}$) of these points."

Page 87

- "The framework includes both 2DCNN and a 3DCNN which are applied depending on the data and tasks reported in the discussions (Chapter 6)" ---> "A 2DCNN and a 3DCNN are implemented for this framework, in order to experiment and understand possible advantages and disadvantages these two may bring in terms of accuracy or computational efficiency, which are discussed in Chapter 6."

Page 88

- Figure 3-6: **"**P_n denotes points**"** ---> **"**P_n denotes number of points**"**
- Added: "The network is schematized with layer parameter settings, aside from the last Dense layer that has the output dimension set to number classes per dataset."
- "Unlike rendering-based or voxel-based methods (Guo et al., 2020), our CNN methods use pseudo images, as shown in Figure 5." ---> "Unlike rendering-based or voxel-based methods (Guo et al., 2020), our CNN methods use pseudo images, as shown in Figure 3-8."
- "The matrix is then sorted by the coordinates, which is observed to provide fractionally better results." ---> "The matrix is then sorted by x- and z- coordinates respectively, which is observed to provide fractionally better results."
- "The 2D matrices are folded along the features' axis (vertical axis in Figure 3-6) in order to adapt the abovementioned 2D matrices for a 3DCNN." ---> "The 2D matrices are reshaped along the features' axis (vertical axis in Figure 3-6) in order to adapt the abovementioned 2D matrices for a 3DCNN."

Page 90

- Added: "For the training of DL models, the F1 score is preferred as a loss function as it is commonly used for assessing the performance of a classifier. However, the F1 score is not suitable as a loss function. In fact, the F1 score is based on counted metrics (TP, TN, FP, FN), which prevents its implementation as a loss function. Therefore, an approximation to F1 score is implemented based on the predicted probabilities, rather than counted classification results. The chosen optimizer is stochastic gradient descent (SGD) due to its performance. The patience is set to 15 epochs for early stopping, observing the validation loss. The learning rate is set to 0.001 and the training is limited to 100 epochs. Besides, in order to handle class imbalance in datasets, the training samples are weighted (i.e., reducing the weight of classes which are represented more in the dataset depending on the occurrences)."

Pages 89, 90, 138

- Figures 3-7, 3-9, and 5-4 are updated with emphasizing the details of "Dense" and "Dropout" layers.

- Accuracy assessment of the classification results are done with the F1 score, overall accuracy (OA), and intersection over union (IoU), along with weighted versions of them with the formulas shown in Table 4-1." ---> "Accuracy assessment of the classification results are done with the F1 score (also known as Sørensen–Dice coefficient), overall accuracy (OA), and intersection over union (IoU, also known as Jaccard index), along with weighted versions of them with the formulas shown in Table 4-1 (Verma and Aggarwal, 2020).

Chapter 4
Page 97
- "… IoU metric represents an area, …" ---> "… IoU metric can be represented as an area …"

Page 98
- Added: "As shown in **Error! Reference source not found.**, IoU represents the ratio of the two areas the overlapped area between the ground truth and the prediction, and the union of these two areas. Scikit-Learn library (Pedregosa et al., 2011) is used for calculation of all the accuracy metrics."

Page 98
- Moved to Page 90: "For the training of DL models, the F1 score is preferred as a loss function as it is commonly used for assessing the performance of a classifier. However, the F1 score is not suitable as a loss function. In fact, the F1 score is based on counted metrics (TP, TN, FP, FN), which prevents its implementation as a loss function. Therefore, an approximation to F1 score is implemented based on the predicted probabilities, rather than counted classification results. The chosen optimizer is stochastic gradient descent (SGD) due to its performance. The patience is set to 15 epochs for early stopping, observing the validation loss. The learning rate is set to 0.001 and the training is limited to 100 epochs. Besides, in order to handle class imbalance in datasets, the training samples are weighted (i.e., reducing the weight of classes which are represented more in the dataset depending on the occurrences)."
- Table 4-1, caption, added: "… $\hat{y}_p$: prediction for *p*-th sample, y_p: ground truth for *p*-th sample, *1(x)*: indicator function"
- Table 4-1, reference added: Scikit-Learn, 2021

Page 100
- "…number of returns, return numbers as well as IR-R-G orthophoto with infrared, red, and green channels." ---> "…the number of returns and return numbers are provided within the classification benchmark dataset. Besides these, IR-R-G orthophotos (infrared, red, green channels) are also provided, which we exploited in our experiments."

Page 101
- "Being large-scale, the dataset is…" ---> "The dataset is…"

Pages 101, 102, 104, 105, 106
- Table 4-3, Table 4-4, Table 4-5, Table 4-6, Table 4-7 "100%" ---> "100.00%"
- Table 4-3, Table 4-4, Table 4-5, Table 4-6, Table 4-7 "Validation" ---> "Evaluation"

Page 107
- "Here, the voxel dimensions are calculated with a leaf coefficient parameter and the original resolution of the point cloud." ---> "Here, the voxel dimensions are calculated by multiplication of a manually set coefficient parameter (will be called as leaf coefficient) and the original resolution of the point cloud."

Page 109
- "… training and validation …" ---> "… training and evaluation …"
- "…includes training, validation, and…" ---> "…includes training, evaluation, and…"

Page 110
- Figure 4-9: "…validation…" ---> "…evaluation…"

Page 113
- Table 4-12 "% of kept points" ---> "Ratio of kept points"
- "The ratio of kept points reported in Table 4-12 differs between 52-94%. This variety of ratios are due to the differences among the original point clouds' resolutions." ---> "The ratio of eliminated points reported in Table 4 12 differs between 52-94%. This variety of ratios are due to the differences among the original point clouds' resolutions."
- "In order to have a correct and fair accuracy assessment, the classification outputs are projected back to the original point cloud." ---> "In order to have a correct and fair accuracy assessment, the classification outputs are projected back to the original point clouds based on nearest neighbors."
- "…the validation datasets …" ---> "…the evaluation datasets …"

Page 114

- Added: “As seen in tables above, DL classifiers achieved higher accuracies (≳80% goal) in terms of F1 scores and OA. The RF classifier failed to achieve 80% OA goal by 1.4%.”

Page 117

- Added: “As seen in tables above, DL classifiers achieve higher OA as in the previous dataset results. However, for this dataset, the OA gaps between the classifiers are less.”

Page 120

- Added: “As seen in the tables above, 2DCNN achieves the highest per-class accuracies, which is also reflected in the OA. Rankings of the per-class accuracies, average F1 scores and OA show similar characteristics to the results represented for ISPR Vaihingen dataset in Table 4-13 and Table 4-14.”

Page 125

- “Instance segmentation results are not shared for this dataset, as this is in a different scale compared to the geospatial point clouds.” ---> “Instance segmentation results are not shared for this dataset, as this is in a different scale compared to the geospatial point clouds and the parameters for instance segmentation is fixed for geospatial data.”

Chapter 5

Page 138

- “The state-of-the-art comparisons include not only point cloud classification frameworks but also an alternative DNN to classify with our data EfficientNet, B7 version specifically (Tan and Le, 2019).” ---> “The state-of-the-art comparisons include not only point cloud classification frameworks but also an alternative CNN to classify using our data matrix structure (Figure 3-6) EfficientNet, B7 version specifically (Tan and Le, 2019).”
- “…EfficientNetB7 cannot be fed with images smaller than…” ---> “…EfficientNetB7 cannot be fed with arrays smaller than…”

Page 139

- “As a result, even with the combination of a mid-level laptop CPU and a decent GPU, both of our DL models are faster than the EfficientNet, while achieving higher accuracies on par with the current state-of-the-art methods.” ---> “As a result, even with the combination of a mid-level laptop CPU and a decent GPU, both of our DL models are faster and more accurate than the EfficientNet, while achieving accuracies on par with the current state-of-the-art methods.”

Page 141

- “…outperform the reference methods by 1–4% in average…” ---> “…outperform the reference methods (PointNet++ and HDA-PointNet++) by 1–4% in average…”.

Page 144

- “…results than RF classifier.” ---> “…results than RF classifier with the feature space defined in Section 3.2.”

Page 146

- “The reported results indicate that the TONIC framework can outperform the current state-of-the-art methods by a few percent of OA, as shown in Table 5-11 and Table 5-14, while requiring less memory and energy consumption due to its design.” ---> “The reported results indicate that the TONIC framework can outperform the current state-of-the-art methods by a few percent of OA (Table 5-14), while requiring less memory and energy consumption due to its design (Table 5-11).”

Page 149

- Added: “Based on these, it can be seen that the objectives of the thesis as mentioned in Section 1.5 are also achieved as the introduced perspective is actualized and validated.”

Page 154

- “this paper” ---> “this thesis”

Appendices A

Page 169

- Added: “The codes shared in this thesis are not meant to be used for any purpose without an official agreement with Skoltech and Bruno Kessler Foundation.”