Skoltech
Skolkovo Institute of Science and Technology

Thesis Changes Log

Name of Candidate: Emre Özdemir

PhD Program: Engineering Systems

Title of Thesis: Geospatial Point Cloud Classification

Supervisor: Prof. Alessandro Golkar, Dr. Fabio Remondino

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

General Corrections:

p.153: "this paper" => "this thesis"

p.145: "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)."

p.138: "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."

p.137: "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)."

p.112: Table 4-12: "% of kept points" => "Ratio of kept points"

p.112: "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."

p.99: Table 4-3, p.101: Table 4-4, p.102: Table 4-5, p.104: Table 4-6: "100%" => "100.00%"

p.97: added: "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 occurancies)."

p.94: 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).

p.87: "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."

p.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).”
p.6: “(ii) Generalization of the learned classification ability (i.e., predicting on datasets with acquisition setup)” => “(ii) Generalization ability (i.e., predicting on datasets with different acquisition setup)”
p.54: “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 (≳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 (≳80%) with a more efficient (in terms of computational time and hardware requirements) methodology.”
p.113: “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.”
p.88: “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.”
p.81: Figure 3-3: Images updated to match with the legend.
p.85: Figure 3-4: Legend added showing color scale low to high values.
p. all pages: Sentences with we/us/our revised and these words removed.
p.59: “An Overview of the Artificial Intelligence” => “An Overview of Artificial Intelligence”.
p.51: Figure 1-12: “… for each planar segment…” => “for each extracted planar segment”.
p.66: “… can achieve high accuracies such as…” => “… can achieve accuracies such as…”.
p.140: “…outperform the reference methods by 1–4% in average…” => “…outperform the reference methods (PointNet++ and HDA-PointNet++) by 1–4% in average…”.
p.143: “…results than RF classifier.” => “results than RF classifier with the feature space defined in Section 3.2.”
p.98: “…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.”
p.43: “There are two commonly used geospatial point cloud types;…” => “There are two common data types representing 3D geospatial data;…”
p.43: "...represents the height. In Figure 1-7,..." => "...represents the height. This procedure is also known as rasterization. In Figure 1-7,..."
p87, p88, p.137: Figure 3-7, Figure 3-9, Figure 5-4 are updated with emphasizing the details of “Dense” and “Dropout” layers.
p.87: 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.”
p.56: "...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."
p.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."
p.88: "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."
p.79: 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.”
p.85:"Local planarity is the average distance between the neighboring points (p_k) to the best-fit plane ($\vec{P}$) of these neighborhood points." => "Local planarity is the mean distance between the neighboring points (p_k) to the best-fit plane ($\vec{P}$) of these points."
p.98: added: “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.”

p.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%."
p.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."
p.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."
p.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."
p.97: 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."
p.1: added: "Associate"