

Answers to the Reviewers Comments

Reviewer 1:

- 1- This paper proposed a classification method for RGB and hyperspectral data based on 3D-CNN model. The results prove 3D-CNN have the capability to revolutionize the species classification tasks.

Answer:

Thanks for your comments to our article. We appreciate the time and effort that you spend to improve this work. We wish that the following answers address all of your concerns.

It has some merits, but also with several problems. Therefore, I suggest to resubmit this paper with following comments in details.

1. In order to better highlight the values of 3D-CNN, please elaborate on the similarities and differences between MLP and CNN. Although CNN originates from MLP, the relationships between the two methods are still unclear in the article.

Answer:

The following lines have been added to 2.3.2 to address this concern:

“CNNs and MLPs are very similar in nature, since they are both frequently employed to estimate highly non-linear functions of any type. MLPs could be employed as a sub-part of CNNs. The main difference between CNNs and MLPs relates to the existence of convolution and pooling in CNNs that plays an existential role to overcome the curse of dimensionality as the data moves through a CNN.”

- 2- Through comparing these experiments, it can be seen that CHM is not useful for 3D-CNN classification. Could you analyze the reason? Do the experiments prove that 3D point clouds cannot support the classification of trees species?

Answer:

The following sentences have been added to 3.results and discussion to address this concern:

“Studies such as [19] has already demonstrated lower classification accuracy for point cloud only features in comparison to a combination of CHM and spectral channels. In [19] the lowest accuracy was related to birch which was relatively indistinguishable from pine. In this regard, our results is consistent with [19].”

- 3- The three major tree species in a boreal forest, spruce and birch own some biological differences. It is pretty important for readers to understand the typical characteristics of classification objects in trees. Add these keys points in the paper. Meanwhile, please give a brief analysis about the main difficulties in classifying tree species based relevant literatures.

Answer: The following lines have been added to answer this concern:

The three main tree species in Finnish boreal forest are pine, spruce and birch. Although the 3D shapes of coniferous pine and spruce crowns are usually clearly different (spruces being more conical [14]), their canopy-level spectra are quite similar [15]. Birch trees, however, have often more distinguishable canopy-level spectra, especially at near-infrared due to higher leaf-level reflectance at NIR and higher contribution of first-order scattering ([15], [16]).

Reviewer2:

This paper studies the usage of 3D convolutional neural networks (3D-CNNs) for classifying different boreal forest species based on hyperspectral (HS), RGB channel, and canopy height model (CHM) data. Different combinations of inputs are considered and used to train CNN models. The accuracy performance of the models is assessed to examine their effects and compared against a multi-layer perceptron (MLP) classifier as a benchmark. The results of the study reveal the effectiveness of using 3D-CNN-based models for the classification problem compared to the MLP, with the best overall accuracy obtained using the 3D-CNN trained leveraging HS and RGB input data.

The subject of this paper is interesting and relevant for the journal, especially considering recent advancements in robotics for environmental monitoring applications such as forestry. The general approach followed to tackle the problem of tree species classification is logical and theoretically sound.

Answer:

Thanks for the time and effort that you spend to improve this work. We carefully went through all of your comments and we wish that the following answers address all of them.

- 1- However, the reviewer's main concern with the paper is that its **exact scientific contributions and novelty are unclear.**

Answer:

The following sentences have been added to address this concern:

In general, we followed a similar path to studies such as [7] in classification methodology. In feature selection, we have tried to complete studies such as [1] and [19]. The main contribution of this work includes:

- 1- proposing an efficient structure for a 3D-CNN network that is suitable for tree species classification; the proposed model outperforms previous studies such as [1], [19],
 - 2- evaluating the proposed model and comparing it to MLP,
 - 3- comparing different combination of features to propose the most effective feature set among the features of this study.
-

- 2- The usage of 3D-CNNs for classification is a well-studied topic and **it is not clear which aspects of the methodology bring new knowledge to this field of study.**

Answer:

This paper is based on our understanding the first work that employs a 3D-CNN in tree classification. Other contributions could be seen in the previous answer as well.

- 3- For instance, is the application of 3-D CNNs with combined HS/RGB/CHM inputs, or the data collection setup, relatively novel or underexplored in previous work? If so, **the authors**

should state this explicitly in the Abstract and Sections 1 and 4 to emphasize the new insight their study brings.

Answer:

This is addressed in introduction part.

- 4- To this end, an explicit statement of the contributions of the paper in the Section 1, perhaps in bullet points, is advised. This would make the implications of the research stronger.

Answer:

The bullet points of the contribution have been added to introduction.

- 5- Importantly, it is also unclear what the exact objective of the paper is; specifically, whether it aims to develop and compare 3D-CNN classifiers with different inputs for this task, or design the best classifier given the data (since "highest ever presented classification accuracies" are reported in Section 4). How does the comparison with the MLP classifier contribute to this research aim? What practical implications do the findings have on real-life data collection scenarios, e.g., sensor setups that can be equipped on UAVs?

Answer:

The objective is stated in last part of the introduction as : "Objectives of this study were to develop and compare 3D-CNN models and features for classification of three major tree species in the boreal forest." Therefore, two main objectives were defined: 1- developing a 3d-cnn model for tree species classification and comparing it to other studies and MLP, 2- comparing different feature sets.

Sections 2 and 3 are generally understandable and well-structured. The theoretical foundations of the approach appear to be sound. The details of the experimental setup are particularly interesting and well-explained.

- 6- However, here, too, the paper would strongly benefit from a clearer statement of its contributions.

Answer:

The contributions have been stated in last part of introduction. restating the contributions seems redundant.

Currently, the theoretical sub-sections, e.g., Section 2.3 describing the classification models, read in a very basic and descriptive manner.

- 7- To improve the focus of the paper, it would help to connect the ideas here to the main objectives of the paper.

Answer:

Line 288-293 in section 2.3.2 clearly describes the process of designing the 3D-CNN for the objective of the paper.

Similarly, in Section 3, the authors are advised to state more explicitly the purpose of the experiments and further discuss the implications of their conclusions.

In general, the article is readable and logically organized. However, there are some issues in wording, grammar, and formatting which are distracting for the reader.

- 8- A proofreading is recommended particularly with regard to: usage of articles, correctness/consistency of notation/typography in equations, acronym usage, sentence structure, consistency of past/present tense, formatting of tables/figures, layout. Some minor suggestions are given below.

Answer:

The article was sent for proofreading.

Overall, this work is an interesting and relevant paper on the topic of tree species classification using 3-D CNNs with different input data. The theoretical foundations of the approach are sound. However, before acceptance, the authors are advised to revise their submission to highlight their contributions more explicitly and improve the focus of the paper.

Minor comments/suggestions.

- 9- p. 1 of 14
 - * Abstract. It is not clear what "image16" refers to.

Answer:

Fixed.

- 10- * Abstract. "... in most of layers." -> in most layers.

Answer:

Fixed.

- 11- * Section 1, par. 1. Opening sentences would benefit from more citations.

Answer:

Done.

- 12- p. 2 of 14
 - * "Each classification technique has its own benefits and drawback." -> sentence is a little too general and does not lead into the research topic well

Answer:

Edited.

13- p. 3 of 14
* "UAV" acronym undefined

Answer:

It is now expanded upon the first usage.

14- p. 4 of 15
* "Totally 3896 trees..." -> A total of 3896 trees

Answer:

Edited.

15- * $g(a) = \tanh(a)$ -> inconsistent italicization

Answer:

Edited.

16- p. 5 of 14
* "... \hat{y} is classification result..." -> is THE classification result

Answer:

Edited.

17- * "... l is number of layers." -> l is THE number of layers

Answer:

Edited.

18- * Figure 2. It might be helpful to point out that this is the network used in the study.

Answer:

This figure (new number is 4) is addressed in line 251 as: The structure of the MLP used in this study is illustrated in Figure 4.

19- * "Convolutional neural networks are related to MLP's." -> MLPs (without apostrophe)

Answer:

Edited.

20- * "CNNs contains two parts: ..." -> contain (without s) two parts

Answer:

Edited.

21- * Section 2.3.2. This section is very descriptive and especially loses focus from the purpose of the experiments (see above).

Answer:

This section is designed to support the CNN concepts that are used in the article. The text is as descriptive and as compact as possible. The section in the authors opinion focuses on its purpose. Please specifically point out to any obligatory changes, if a modification is required.

22- p. 6 of 14
* Revise English - fragmented and unclear

Answer:

A revise is done.

23- * ReLu vs. ReLU acronym

Answer:

All "ReLU" edited.

24-
* Lines 245-253. Description of the model training procedure could be moved to Section 3.

Answer:

The paragraph is moved to 3.2.

25- * Section 2.4. It would be helpful to restate the purpose of the experiments here to connect it to the performance metrics.

Answer:

The initial sentence is changed with: "The purpose of this study was to investigate applicability of 3D-CNNs to classify tree species from HSIs, therefore, evaluating the suggested network models is an essential part of the study. Here we used overall classification accuracy to measure the performance of a given model as"

26- * Section 2.4. Equation numbers seem to reset.

Answer:

Edited.

27- p. 7-8 of 14

* Lines 271-279 simply repeat information from the table. This could be replaced with more interpretation of the results.

Answer:

Edited.

28- p. 9 of 14

* "... it seems that the CHM layer did not improve..." -> DOES not improve

Answer:

Edited.

29- * Could the results be explained intuitively by looking at the structure of the trees? Perhaps it would be interesting to show example classification results.

Answer:

The discussion and conclusion is edited to address this concern.

30- * Table 4. It would be useful to highlight the main results in bold for clarity.

Answer:

The paragraphs are edited to address this concern.

31- p. 10 of 14

* Lines 349-350. This is a strong statement and it is unclear how it validates the aims of the study.

Answer:

The classification metrics on an independent dataset support the claim of this sentence, since both datasets were similar between these two studies.

32- p. 12 of 14

* Section 4. It is unnecessary to restate the entire experimental procedure, only summarize it and the main findings.

Answer:

Conclusion is edited to address this concern.

33- * Lines 386-388. Strong statement - relative to which methods?

Answer:

The conclusion is edited to address this concern.

34- * Line 383. How is the "proposed model" new/innovative wrt. previous work

Answer: We discussed the novelty of the proposed model in two aspects: 1- classification accuracy, 2- simplicity. The proposed model outperformed the best models while being simpler (98% less parameters) comparing to the recently proposed models by other researchers.

Reviewer 3:

the manuscript describes a number of deep-learning applications (CNN) fusing different combinations of sensors from UAV. The paper is well written, highly significant, and useful to move the field forward. I particularly enjoyed the comparison of different combinations of RGB, HSI, and CHM. I also enjoyed the long and well written description of neural networks in section 2.3.

Answer:

Thanks for your encouraging comment to our article. We appreciate the time and effort that you spend to improve this work. We wish that the following answers address all of your comments.

Before acceptance, there are some minor edits that I would suggest:

- 1- The authors found that CHM contribution to classification is generally low or negligible. Despite I can confirm we had similar results from a similar application, this is potentially counter-intuitive to the readers. For that reason I would enjoy having a small paragraph discussing why the authors believe fusion of HSI and CHM fails in their application.

Answer:

To address this comment a new sentence has been added.

- 2- line 101: appears to me that 11nm (and not um) is the band-width, and not the pixel size. This would make sense with how authors aggregated HSI bands (line 107). Am I right?

Answer:

We didn't find any "um" unit. "nm" is used instead in all occasions.

- 3- Table 3 could probably be dropped in Supplementary materials

Answer:

It is moved to the supplementary materials.

- 4- line 277: despite many authors report accuracy for training set, I personally would rather drop them, because they could be easily misinterpreted by less expert users (but I acknowledge the fact that this is just personal taste).

Answer:

We totally agree that reporting classification accuracy on the training set could be misleading. For this reason, we considered a separate validation and test dataset to stop the training through an appropriate mechanism and independently evaluate our classifier.

5- In line 308 and later the authors use the term "demonstrate", but I'd personally prefer using "show"

Answer:

An edit is applied for many occasions.

Nice work!

Answer:

Thanks!