Manuscript Number: CEUS-D-23-01167

Decoding (urban) form and function using spatially explicit deep learning

Dear Dr Fleischmann,

Thank you for submitting your manuscript to Computers, Environment and Urban Systems.

I have completed my evaluation of your manuscript. The reviewers recommend reconsideration of your manuscript following major revision. I invite you to resubmit your manuscript after addressing the comments below. Please resubmit your revised manuscript by Mar 31, 2024.

When revising your manuscript, please consider all issues mentioned in the reviewers' comments carefully: please outline every change made in response to their comments and provide suitable rebuttals for any comments not addressed. Please note that your revised submission may need to be re-reviewed.

To submit your revised manuscript, please log in as an author at https://www.editorialmanager.com/ceus/, and navigate to the "Submissions Needing Revision" folder.

Computers, Environment and Urban Systems values your contribution and I look forward to receiving your revised manuscript.

Kind regards,

Yang Yue, Ph.D.

Associate Editor

Computers, Environment and Urban Systems

*Thank you so much for the opportunity to let us resubmit our paper. We have included detailed descriptions and responses to the original comments and flagged where and how the original manuscript has changed.*

Editor and Reviewer comments:

Reviewer #1: The main contribution of the manuscript lies in utilizing deep learning and satellite imagery to understand urban landscapes. I think this idea can have application value in urban planning and environmental management. The manuscript explores a new perspective for analyzing urban landscapes by combining deep learning with spatially explicit methods, integrating a profound understanding of spatial characteristics, thereby enabling more effective understanding and classification of urban forms and functions. However, the manuscript also has some limitations, such as a lack of logical coherence in the narrative, unclear descriptions of methods, and insufficiently detailed and in-depth analysis of results. I have spent lots of time to understand the contribution of the manuscript , since spatially explicit deep learning is not new in urban study. So I think the author(s) need to revise the manuscript carefully in the (possibly) next round.  
  
1.The abstract should be improved to clearly summarize the methods, experiments, and results of the research. While this abstract mentions the use of deep learning methods and satellite imagery to decode urban landscapes, it does not detail the specific methods employed (e.g., the type of model), the design of the experiment (like comparative models), or the main results (e.g., the accuracy of the model's testing results). These details are crucial for readers to understand the contributions of the manuscript.

*Thank you for pointing out the lack of clarity of the original abstract. We have completely rewritten the abstract in the revised version of the manuscript and believe that its current version matches the expectations.*  
  
2.The manuscript indicates that the most part of the deep learning methods recently introduced into urban satellite imagery analysis overlooked the geographical characteristics of the images they process. What exactly do these geographical characteristics refer to? What role do they play in the research objectives of this manuscript? What consequences will appear if they are ignored? The author needs to provide a detailed explanation.

*Thank you for the recommendation. We agree that the original description was a bit too succinct. We have included a dedicated paragraph in the revised version of the introduction answering the questions above and providing further explanation of what is meant by missing geographical aspects.*   
  
3.The "spatially explicit method" is a key aspect of the methodology, but the introduction lacks a detailed explanation to this method and its relevant applications. The author needs to substantiate the advantages of this method through related literature and explain how it will be applied.

*We have expanded the introduction to include an additional paragraph containing deeper explanations of the what is meant by spatially explicit methods and how can they be introduced into the predictive pipeline based on neural networks.*  
  
4.The introduction section overall lacks coherence. The author should clearly itemize the contributions and innovations of the manuscript, including any novel methodologies, unique data processing techniques, or significant improvements to existing technologies. Additionally, the summary of the contents of each chapter in the final part of the introduction is quite vague, making it difficult for readers to capture information from it.

*Thank you for pointing out the lack of clarity and structure in the introduction. In the revised version of the manuscript, we have restructured and significantly expanded the section and including explicit itemisation of the main contributions of the manuscript. Furthermore, we have expanded the last paragraph to include specific description of the remaining contents of the paper.*  
  
5.The manuscript lists typical and authoritative land classification results such as CORINE, the European Space Agency's WorldCover project, and Esri's land cover classification, however, it merely describes them briefly and directly concludes that these classifications are not very suitable for research applications focusing on urban environments. It is recommended that the author provide a comparative analysis of why these methods were abandoned in favor of spatial feature classification methods. Additionally, presenting this comparison in a visual format, such as tables or charts, would provide a more intuitive help in understanding the rationale behind choosing a different approach.

*Thank you for pointing out the lack of clarity of reasoning here. We have now included a visual comparison showing spatial distribution of classes from three established LULC products alongside the spatial signatures to ensure the reasons motivating us to use latter rather that former is properly understood by a reader. We believe that the visual and its caption provide additional context needed to understand the rationale behind the selection. While we would love to do a proper comparative analysis aimed at understanding the feasibility of these LULC data within the proposed framework, we believe that a work of such extent shall be within an independent paper and hence consider it out of scope (and primarily out of space) here. We do acknowledge this research avenue in the discussion.*  
  
6.One of the goals of sampling chips with specific size is to find a balance between the number of chips sampled from the data and the amount of information each chip can hold. How is the goal specifically achieved? Or has it been achieved? If so, please provide relevant explanations.

*The first paragraph of the Discussion section in the original manuscript summarized the outcome of the analysis leading to an assumption than larger chips may be better. Though this cannot be interpolated. Based on your helpful suggestion, we have expanded the relevant section to add an explicit conclusion on the question of size x data question.*  
  
7.In Section 2.2 Methods, it is mentioned that image segmentation predicting individual pixels would not allow us to use the second part of this method and effectively test the impact of spatial lag in modeling. What does the second part of the method specifically refer to? What impact does spatial lag have in method modeling? What is the purpose of testing the impact of spatial lag in modeling? The description of these contents can be also improved.

*We have rephrased the section of the text referring to the issue of using segmentation within the modelling as the original one was slightly confusing. The main point we wanted to raise is that the part of the pipeline based on logit ensembles and histogram-based gradient-boosted trees using the spatial lag would be extremely computationally expensive. We believe that there is a higher value in having a reproducible method than one depending on a high-performance cluster. Given this is not the only reason for exclusion of segmentation, we believe that there is no need to verify such assumptions empirically. Given the expanded introduction providing deeper insight into the reasoning of why spatial lag is relevant, we believe that the reworked section is now clear.*

8.The manuscript lacks a clear model architecture diagram. It is recommended that the author add a complete model framework diagram for the proposed method to intuitively display the composition of the competitive architecture and the functions of each part.

*Thank you for the suggestion. We agree that the visual illustration of the model architecture help the understanding and we have added one as a Figure 5 of the manuscript.*  
  
9.The experiments in this manuscript are primarily focused on specific target data represented by spatial features, and the behavior of the spatial components in the modeling pipeline may differ when target data is of a different nature. The author believes that this principle still holds true in most cases, but this viewpoint lacks the necessary proof and is not convincing.

*We have further expanded the relevant section, adding further reasoning why we believe this is the case and specifying in which situation we believe it holds. We also acknowledge that this is only a hypothesis until tested on other data. However, such an empirical test is out of scope of this article.*  
  
10.The manuscript needs further strengthening in explaining the experimental results and their potential applications. On the one hand, the author has conducted relevant analyses from four performance metrics, but for final target—advance our understanding of the extent to which conventional deep learning methods can be applied to satellite imagery to capture the composition of urbanised landscapes, these analyses do not intuitively explain the direct value of the experiments to the main idea. On the other hand, the manuscript should also explore how the model can be specifically applied in the fields of urban planning and environmental management, such as identifying patterns of urban expansion, planning new urban infrastructure, or assessing urban ecosystem services.

*Thank you for the suggestions. We have included an additional paragraph in the Discussion, covering the explicit answer to the question of understanding whether NNs can be used within the set context, which was only implicitly provided in the original text. The same paragraph now also includes a brief discussion of applicability of the models but given the technical nature of the paper, we kept the section succinct.*   
  
  
Reviewer #2: The manuscript introduces a deep learning-based method for decoding urban function and form information from remote sensing imagery. However, there are several areas that require further clarification and improvement.  
  
Major issues:  
1. It would be beneficial to include graphical illustrations when introducing the concept of spatial signature and basic spatial unit of ETC, as well as providing a brief introduction of the different types.

*Thank you for the feedback. We agree that given the novelty of signatures, we could have included more extensive explanation of the concept. We have now expanded the description and added a diagram showcasing the workflow of signature delineation, including the creation of ETCs. We have further included interpretative pen portraits of all signature types as an appendix. Additionally, Figure 2 now shows signatures delineated in one more area (London), in addition to Scottish Belt present in the original version.*

2. The purpose of the new classification system should be clearly articulated, particularly in relation to its applicability to different regions and countries. Additionally, a comparison to the concept of "urban functional zones" should be included in the literature review, which is a popular concept (also widely studied) that try to extend the urban land details of traditional land use and land cover (LULC) classification and seems closely related to the concept of spatial signature.

*Thank you for your recommendations. We acknowledge that introduction of spatial signatures may not have provided all the necessary context and we have expanded the relevant section. We have furhter included a short section on urban functional zones in the relevant section describing British Spatial Signatures. We believe that given this manuscript builds on a published data product that comes with its own paper (Fleischmann and Arribas-Bel 2022) and on a concept of signatures proposed in Arribas-Bel and Fleischmann (2022), further comparison of signatures and UFZs is out of scope of this paper.*

3. In Section 2.2.2, the claim that normal image data augmentation approaches do more harm than good should be supported by experiments. For example, rotation augmentation is quite useful for object detection in remote sensing images. The doubt of the effectiveness of the so-called spatial data augmentation can also be verified in Table 2 where SIC results are significantly lower than that of BIC for most cases.

*We have rephrased the section trying to provide more clarity into our reasoning. While we agree that it could be potentially interesting to have empirical data testing standard augmentation methods against all the other options, the matrix of experiments is already very complex. Given the main contribution of the manuscript lies elsewhere, we believe that inclusion of another option would only hinder understandability of the results and have decided not to perform additional experiments. We agree that the value of SIC is not as high as we have originally anticipated and have acknowledged that in the discussion section.*

4. Chip size and sliding strategy are not new and widely used and discussed in deep learning-based remote sensing research. The significance of chip size and sliding strategy in the context of this research should be explicitly stated.

*We have expanded the discussion section related to the comparison of BIC, SIC and MOR models, explicitly stating the effect of sliding and its significance.*

5. More details are needed to clarify the method of splitting data, whether it is done randomly or by geographical regions.

*The description of the splitting method is now present in the appendix and linked from the main text.*

6. The justification for the setting of hyperparameters for the CNN models should be provided.

*The comparison of the different hyperparameters and their effect on the global performance is included as Appendix A. We have further expanded the section Model architecture in the Methods section to explicitly explain the process behind the selection of top layer architecture and a number of its neuron as well as the learning rate.*

7. Subsection titles should be added to the results section for clarity.

*We have included additional subsections in the Results.*

8. An overall workflow chart at the beginning of the method would enhance the readability of the manuscript.

*Thank you for the suggestion. We agree that the visual illustration of the model architecture help the understanding and we have added one as a Figure 5 of the manuscript.*

Minor issues:  
1. There appears to be an irrelevant comment left on page 13, line 50, during the editing process.

*Thank you for noticing. The comment has been resolved and removed.*