Manuscript Number: CEUS-D-21-00877

Simulating large-scale urban land-use patterns and dynamics using the U-Net deep learning architecture

Dear Mr. Wang,

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 Jul 06, 2022.

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,      
Alison Heppenstall     
Associate Editor

Computers, Environment and Urban Systems

**Reviewer #1:**

The paper reports the result of a study using deep learning to model urban land development by applying the U-Net architecture to a study area of the North China Plain. The subject addressed in this article is worthy of investigation and appropriate to the journal.

Thanks for confirming the value of this study.

However, the methodology is not clearly described w.r.t the data used for the development of the two U-Net models and the novel methodological contributions.

To clarify the data used to train the U-Net models, we added a table in the “Data processing” section.

Table 3. The data used to train U-Net models. (This table is inserted in the “2.4. Data preprocessing”)

|  |  |  |  |
| --- | --- | --- | --- |
| Data | Time | Source | Resolution |
| Urban land use | 1994, 2006, 2018 | Wang et al. (2021b) | 30 m |
| Elevation | 2000\* | Shuttle Radar Topography Mission | 30 m |
| Slope | 2000\* | Derived from elevation | 30 m |

\* Note, the elevation and slope data were used in 1994, 2006, and 2018 despite their being acquired in 2000.

The novel methodological contributions, compared to CA models, of deep learning architecture (U-Net) in simulating urban land-use mainly lie in 1) reducing subjectivity upon tweaking model weights/parameters, and 2) introducing computer vision technologies to “see” urban growth (i.e., sensitive to spatial features) rather than applying regression in a neighborhood scop to compose driving factors into transition rules.

That being said, the U-Net itself is not an innovative architecture (cutting-edge perspective) in the deep learning field. So, we focus the innovation of U-Net on how it could add up to the current CA-mainstreamed urban simulation study: 1) automated to extract the optimum transition rules and 2) extract large scale (256 \* 256 pixels window compared to ~ 10\*10 window of CA models) spatial features into the simulation. However, it is not appropriate to discuss novel methodological contributions in the methods section. Instead, we discussed how the U-Net could contribute to urban simulation studies in the discussion section (4.1. Mimicking real-world urban patterns and dynamics, and 4.2. Reducing subjectivity by automatically constructing transition rules).

Somehow, the conclusions are not well supported by the study results, e.g., "The U-Net model successfully captured neighborhood effects, gravity effects, and linear expansion along transportation routes in the urban dynamics", where "linear expansion along transportation routes" seems to be concluded based on visual inspections of 11 smaller areas.

We agree that the “neighborhood effects, gravity effects, and linear expansion effects” could be abstracted by more appropriate metrics in theory, but current landscape indexes have limited power to reflect these characteristics (Uuemaa et al. 2009; Frazier and Kedron 2017; Pontius et al. 2008). Indeed, assessing the quality of output features remains a significant challenge even in the deep learning community because of the versatile traits in spatial signals that are beyond the mathematical description of simple shape indexes (Wang et al. 2020; Ghosh et al. 2020): it is common to evaluate the performances of deep learning models by observing the resulted images (Gonog and Zhou 2019; Xu et al. 2018). However, we still computed the landscape shape index for readers so that they comprehend the performance of the U-Net simulation with a basic estimation.

In this study, we deliberately illustrated 11 small-scale (1km – 10km) regions to present that U-Net has learned some inherent patterns (i.e., the neighborhood, gravity, and linear effects) of urban development. The reason for choosing small-scale regions is that urban expansion simulation usually performs less accurately in small areas than in larger regions (Pontius et al. 2008). If the simulation in a small area looks sensible, readers are more likely to be convinced that the deep learning architecture could simulate the urban development well on the overall scale.

To explain to readers about our visual inspection-based pattern evaluation, we added the description of “The accuracy metrics and spatial pattern metrics have limited power to fully abstract the spatial patterns in urban land-use simulation (Uuemaa et al. 2009; Frazier and Kedron 2017; Pontius et al. 2008)” in the “4.1. Mimicking real-world urban patterns and dynamics” section, and added “To assess the U-Net’s ability in simulating urban land-use, we randomly selected 11 evenly distributed regions in the research area and visually inspect the resulted maps focusing on if U-Net had captured the important urban development characteristics such as gravity, neighborhood, and linear effects.” in the “2.7 validation and accuracy assessment” section.

The highlight states "U-Net achieves similar accuracies with CA models with lengthen calibration". What do you mean about "lengthen calibration"? How does your study result support "similar accuracies"? What makes the U-Net models better than CA models? Perhaps, the highlight needs to be revised and a corresponding quantitative comparison and discussion are needed in the discussion section.

There are challenges to compare model performances across studies. According to Pontius et al. (2008), two factors have profound influences on the prediction accuracies: 1) The net urban expansion area and 2) the spatial resolution. A positive relationship exists between the figure of merit (FoM, i.e., the prediction accuracy) versus observed net change. Prediction errors vanished when the simulation maps were resampled into coarser resolution. Therefore, the same/close historical urbanization magnitude and spatial resolution is the premise to compare U-Net with other CA models.

In this study, we selected two CA-based studies for comparison because they were conducted in China (similar historical urbanization magnitude) and have the same 30m spatial resolution as our study. Wang et al. (2021c) used a particle swarm optimization algorithm, iterated a range of parameters settings such as self-recognition component, social component, inertial weights, and the number of particles to finally get the best model with an FoM of 0.193 in Zhuji, China. Wang et al. (2021a) incorporated four periods of historical urban land-use as momentums, adopted five models on top of each historical period to develop 17 sub-models, and tested eight different neighborhood sizes (5\*5 to 41\*41) for each sub-model to finally determined the best simulation that achieved an FoM of 0.219 in Beijing, China.

The FoM in this study, computed from 76 prefectures, with the 1st quantile to be 0.177, the 3rd quantile of 0.215, and the median of 0.197, was close to 0.193 reported by Wang et al. (2021c), but slightly lower than 0.219 found by Wang et al. (2021a). The lower FoM of our study is because Beijing is one of the most rapidly urbanized regions in China and a fast urbanization region is more likely to lead to a higher FoM in urban land-use simulation (Pontius et al. 2008). Therefore, U-Net achieves similar accuracies to those CA-based models underthought a length calibration process.

We have rewritten the “4.3. Accurate prediction and robustness in capturing spatial patterns” in the discussion section emphasizing 1) the difficulty in cross-study comparison (effects of urbanization rate and spatial resolution), 2) how we selected the studies to be compared, and 3) the close FoM value between our study and other length calibrated studies.

One of my main concerns is the description of the data used in the study, which is not clear and consistent considering the following listed example lines. You actually used land use maps of 1994, 2006, and 2018 only, plus elevation/slope data (not sure if also of these three years), but the way the data are described is confusing.

We added table 3 in the “Data processing” section to make it clear to readers that only land-use maps (1994, 2006, and 2018) and elevation/slope data were used in this study.

Line 97-98: "used high-accuracy maps of urban land-use between 1993 and 2012 to train the model" and tested the model by the projected spatial distribution of urbanization for 2018

We change “between 1993 and 2012” to “between 1994 and 2006.”

To avoid confusion about how the U-Net was validated using a reference map, we modified the description to “We then projected the spatial distribution of urbanization for 2018 using this model and thoroughly tested its ability to accurately capture the high-level spatial features, shapes, and patterns of urban development against the reference urban land-use of 2018.”

Line 100-101: THIS trained model was then used to project patterns in 2030

To avoid confusion between the two models trained in this study, we modified the description to “Finally, we trained another U-Net model on land-use maps of 2006 and 2018 to project patterns in 2030 based on historical urbanization rates.”

Line 108-109: "We used Landsat data to map urban development in the study area for the years 1994, 2006, and 2018" - I think you are talking about "map land use"

The terms used throughout the manuscript need to be clearly defined or understood and consistent, e.g., urban dynamic map, urban map, urban land-use map, urban land use map, urban pixel number, urban pixels, urban images

We have changed “urban dynamic map, urban map, etc.” to “urban land-use map” throughout the manuscript.

Line 111-113: here you start talking about using different years of data for two different U-net models.

In our first submission, we denoted the different historical periods with different names (1994–2006 and 2006–2018 are referred to as “base,” 2018 and 2030 are called “target,” and the true 2018 urban map is called “reference”) and described the U-Net as “U-Net trained on base/target years.” Such notation could confuse the readers because they need to roll back to the method section to discern the differences between “U-Nets trained on base/target years.”

To emphasize the differences between the two U-Net models, we added a table in the introduction section and append a suffix to them. We deleted the description of training phases (i.e., 1994–2006 and 2006–2018 are referred to as “base,” etc.), and used the “U-Net-A/B”, rather than “U-Net trained on base/target years”, to identify the deep learning model through the manuscript.

Table 1. The U-Net developed in this study. (This table is inserted above the workflow chart)

|  |  |  |
| --- | --- | --- |
| Model name | Training years | Project year |
| U-Net-A | 1994, 2006 | 2018 |
| U-Net-B | 2006, 2018 | 2030 |

Line 196-199: what are the spatial resolutions of land-use, elevation, and slope data, assuming they are rasters?

We added a table 3 in the “Data processing” section to make it clear to readers that all data used were in raster format with a spatial resolution of 30m.

Line 213-215: what's urban map here? Here, what data were used to train which model?

We change “urban map for 1994” to “urban land-use map for 1994.”

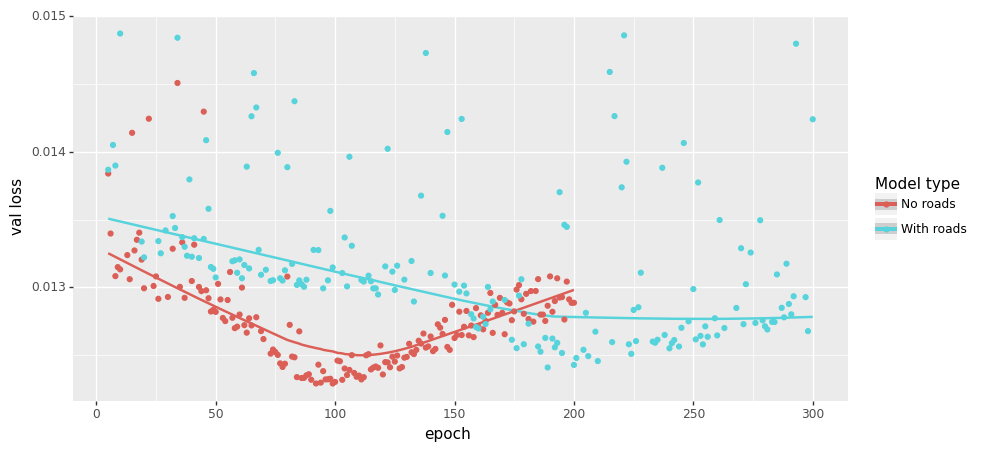
To clarify the data used to train U-Net, we added more information in the parenesis for explanation: “The samples included three-layer input data (e.g., urban land-use map for 1994, elevation and slope for U-Net-A) and a single-layer future urban land-use map (e.g., urban land-use map for 2006 for U-Net-A).”

Line 226: what are the input images here? Land use maps?

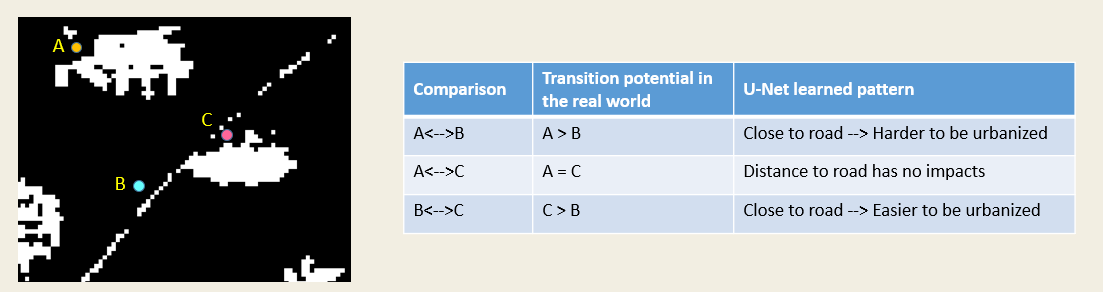
The input images refer to the combination of the land-use map and the elevation/slope data. To clarify this concept, we added more information in the parenthesis for explanation: “The input image (e.g., a three-layer image of urban land-use for 1994 and the elevation/slope data for U-Net-A) was split into tiles, supplied to the trained U-Net model to produce separate outputs, and then the outputs were mosaicked into a single transition potential map.”

Line 204-205: any previous studies that prove "little additional useful information" provided by distance factors?

We had conducted a separate experiment that showed that “distance to road provided little additional useful information.” We collected historical road data (vector) from *OpenStreetMap*, computed the distance to roads (meter records in raster format), and append these distances data as additional drivers to train U-Net. The evaluation (MSE of 5k 256\*256 tiled images against their corresponded reference 256\*256 tiled images) shows that introducing road distance information makes the U-Net harder to train (longer training time and slightly worse MSE).



One possible explanation is that U-Net is sensitive to “visual signals (i.e., spatial features)” but distance hardly provides any robust spatial features. Below is the transition potential comparison of three sites given distance to road. Three contradictory conclusions can be concluded by the U-Net with the road information appended. So, we determined not to add distance information into the U-Net model.



Another commonly used distance information is the distance to the existing urban centers. We also determined not to use such data because it is also against the setting of deep learning for urban simulation: use visual signals to “see” the urban development. Whereas distance information falls under a different principle of “use regressions to capture the first law of geography.” In the end, “to see urban development” is the most exciting part that can be added to the CA-mainstreamed land-use simulation studies, so we had not used any distances data in this study.

To make this claim clear to readers, we provided a supplementary document to explain the reason for not including distances data in our study.

Line 285-286: "to historical urban land-use maps for the period 1994-2006" - the expression implies maps of more than two years

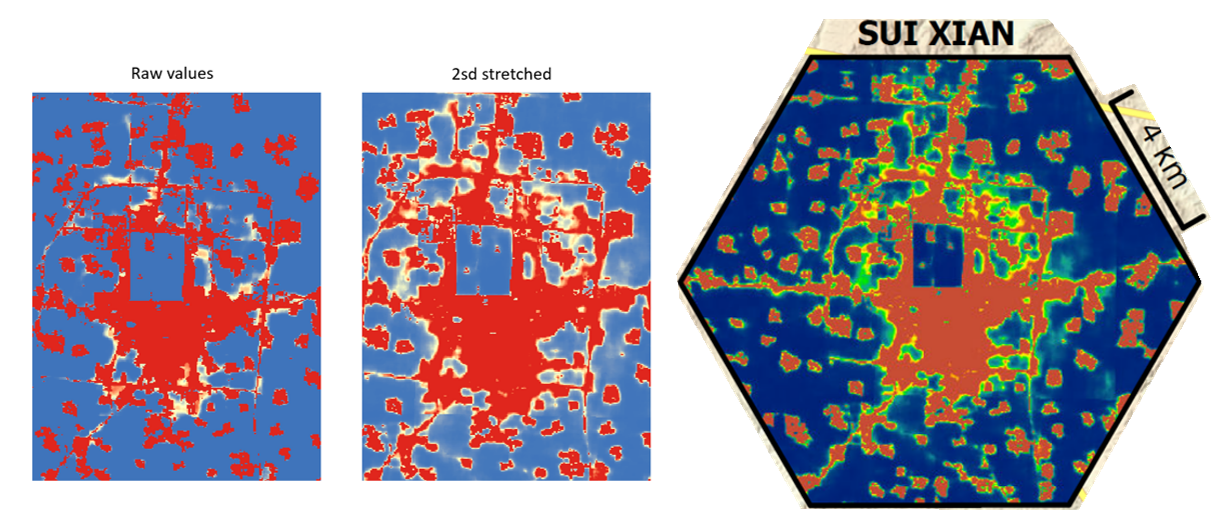
We had changed the description to “historical urban land-use maps of 1994 and 2006”

What message does Fig. 5 convey? How does it show the dynamics of urban land use? What do you mean about "urban dynamic map"? Do you really need this large map while showing the details as small snapshots?

The purpose of Fig. 5 is to introduce the historical urban land-use to the reader. This message connives fewer connections to the main topic of this study “deep learning to simulation urban land-use,” so we deleted this figure.

Line 310-312: Is SuiXian a large town? What are those smaller villages which have the same phenomena - high transition potential?

There are several small towns, including SuiXian, been zoomed into as snapshots of transition potential maps. The visually high transition potential effects are because of a “standard deviation” stretch method being applied to the map. Otherwise, there will be only very thin peripheral areas indicated as high transition potential.



To clarify the visual effects of the transition potential map, we added the stretch method and parameter to the figure note: “The transition potential values were stretched by two standard deviations to better show the learned patterns of U-Net.”

Figure 10 - do you really need this large map while showing the details as small snapshots?

Figure 11 - do you really need this large map while showing the details as small snapshots?

We do need these two figures to demonstrate that U-Net had learned the “gravity, neighborhood, and linear effects” in urban development. We chose to keep the large map as well as show the detailed small inset maps for two reasons: one reason is to give readers the overall impression of the whole research area on how U-Net performed. The other reason is that eliminating the large map leads to less aesthetic attractive designs and won’t reduce much page space (see below our alternative design).

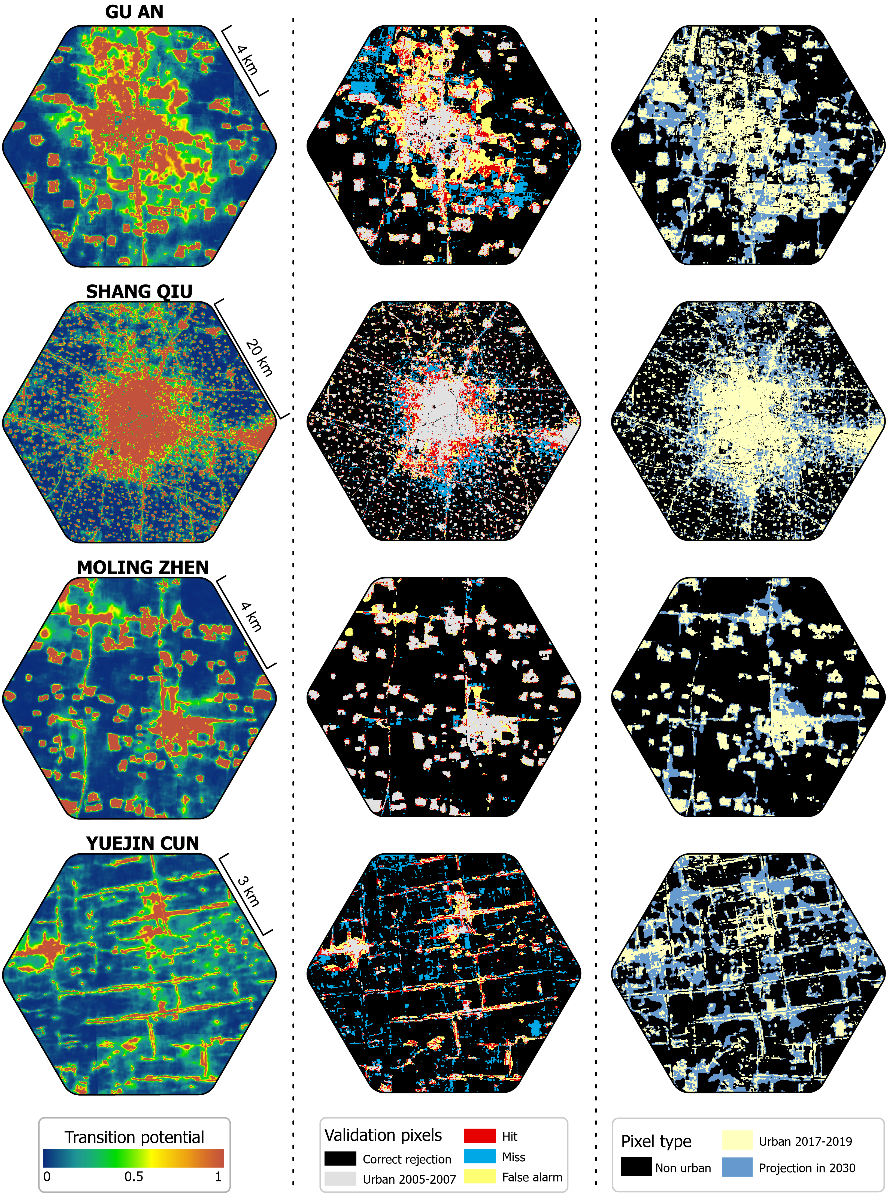
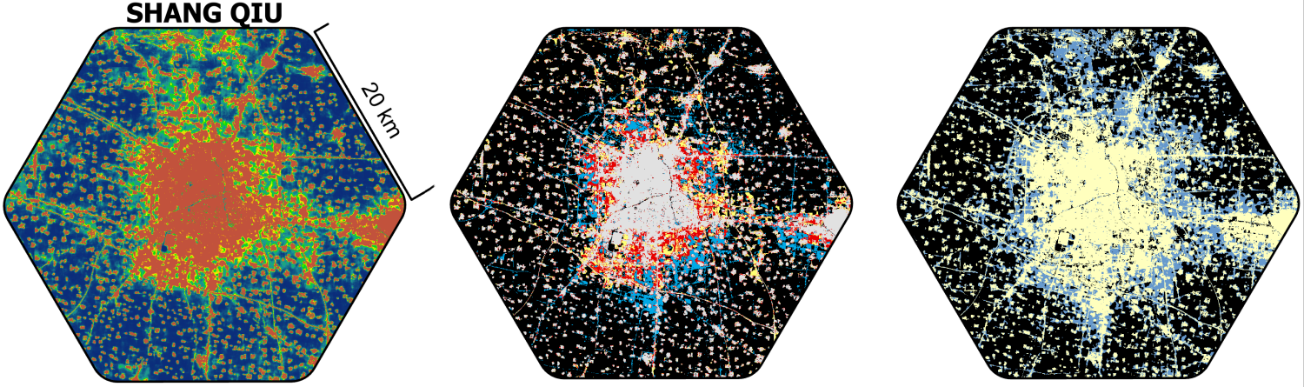


Fig. 10, 11, and 13 show 11 smaller urban areas, no large urban centers. How will you generalize the observations from the patterns around these areas to more general urban development?

Readers can zoom into the large cities to see the generalized simulation of urban land-use with our high-resolution figure (I think there would be a download link pop put near the figure with the online version). We have tried to include large cities as inset maps, but the visual effects are less clear than in the small towns because the subtle details are obscured by large-scale transition potential patterns (see below figure).



What's the purpose of including the projection of 2030 urban land use? Is it just to demonstrate the use of the trained U-Net models?

The complete flow of urban land-use simulation in this study is: 1) develop the U-Net architecture, 2) validate the trained U-Net against historical reference urban land-use map, and 3) apply the U-Net to project urban land-use to a future date. The primary purpose of including the projection to 2030 is to demonstrate the use of the trained U-Net. However, as one of the fastest urbanized regions in the world as well as the food bowl of China, knowing the accurate future urban land-use in the future provides significant practical handles to regulate urban development and evaluate the food sustainability of China. Urban land-use is a key proxy for socioeconomic assessment and environmental protection. Studies focusing on sustainable development and climate change can benefit from our future land-use predictions.

**Reviewer #2:**

I have reviewed this article carefully; I agree that it is a promising new idea to apply U-Net to urban growth simulation and prediction.

Thanks for confirming the value of this study.

However, this method focuses on the characteristics of images (i.e., urban land use landscape), rather than understanding the process of urban growth from the driving factors (i.e., the combined effects of nature and society, etc.) that drive landscape evolution. Therefore, this method cannot understand the mechanism of urban evolution.

There are two types of CA-based urban land-use simulation models: the one is called the regressive model which focuses on increasing the simulation accuracy and finding the optimum functions to compose drivers (e.g., suitability, proximity, stochastic factors, etc.) into transition rules (Gantumur et al. 2020; Shafizadeh-Moghadam et al. 2017). The other one, often called the participatory model, emphasizes balancing the interests of state-holders and materializing different appeals in the urbanization process to reveal the mechanisms/drivers behind the urban land-use change (Clarke and Johnson 2020; Mansour et al. 2020).

The U-Net shares the same goal with the regressive model of increasing prediction accuracies but possesses several significant advantages: 1) automatically extracts the transition rule without manually setting parameters such as neighborhood size and decay rates etc (Feng and Tong 2020; Roodposhti et al. 2020; Peng et al. 2020), and 2) use computer vision technologies to “see” (i.e., sensitive to spatial features) urban development rather than build a regression between driving factors and urban land-use. Therefore, introducing deep learning architectures bring about the significant potential to assist CA-based models with better simulation performances.

That being said, there are several potential ways to transform the trained U-Net into human recognizable rules, i.e., understanding the mechanism of urban evolution. The explainable machine learning techniques enable the abstracting of knowledge from deep learning architectures (Gunning et al. 2019; Dosilovic et al.). However, there are uncertainties if these measures can be successfully applied to U-Net and they also exceeded the topic of this study. To make readers clear about the limited capability of U-Net models in explaining the learned rules as well as the potential ways to extract human-recognizable rules, we modified the discussion on the “limitation section” from “U-Net have limited explanation power” to “U-Net have limited explanation power but there are potential ways to extract learned rules” to entice future studies.

On the other hand, the author does not strongly prove that the simulation accuracy of this method is better than the results of traditional CA-based results. I noticed that the author mentioned in the article that U-NET achieved higher accuracy than CA in some cities. But this does not mean that U-NET performs better in a large area, such as the study area of this paper.

The primary goal of introducing deep learning to urban land-use simulation is to explore how deep learning could add up to, rather demonstrate deep learning is better than, CA-based land-use modeling. The comparison between U-Net and other studies showed that these two ways, from the accuracy metrics perspective, are very close.

The accuracy of the simulated urban land-use is profoundly impacted by 1) the net urbanized area and 2) the spatial resolution of the raster data (Pontius et al. 2008). We compared our result with two CA-based studies that iterated multiple parameters (window sizes, inertia weights, iteration number, etc.) or developed 17 submodels to get the best simulation outcome (Wang et al. 2021a; Wang et al. 2021c). The comparison showed that the FoM in this study, computed from 76 prefectures, with the 1st quantile value of 0.177, the 3rd quantile of 0.215, and the median of 0.197, was close to 0.193 reported by Wang et al. (2021c), and 0.219 found by Wang et al. (2021a).

U-NET achieved higher accuracy than CA in some cities because the FoM is positively related to the urbanization rate (Pontius et al. 2008): some cities had more net urbanized areas and are more likely to have higher FoM. However, the interquartile value of FoM from U-Net simulation is very close to the selected CA-based studies. Therefore, U-Net achieves similar accuracies to those CA-based models underthought a length calibration process.

Publication bibliography

Clarke, Keith C.; Johnson, J. Michael (2020): Calibrating SLEUTH with big data: Projecting California's land use to 2100. In *Computers, Environment and Urban Systems* 83, p. 101525. DOI: 10.1016/j.compenvurbsys.2020.101525.

Dosilovic, Filip Karlo; Brcic, Mario; Hlupic, Nikica: Explainable artificial intelligence: A survey, pp. 210–215. DOI: 10.23919/MIPRO.2018.8400040.

Feng, Yongjiu; Tong, Xiaohua (2020): A new cellular automata framework of urban growth modeling by incorporating statistical and heuristic methods. In *International Journal of Geographical Information Science* 34 (1), pp. 74–97. DOI: 10.1080/13658816.2019.1648813.

Frazier, Amy E.; Kedron, Peter (2017): Landscape Metrics: Past Progress and Future Directions. In *Curr Landscape Ecol Rep* 2 (3), pp. 63–72. DOI: 10.1007/s40823-017-0026-0.

Gantumur, Byambakhuu; Wu, Falin; Vandansambuu, Battsengel; Tsegmid, Bazarkhand; Dalaibaatar, Enkhjargal; Zhao, Yan (2020): Spatiotemporal dynamics of urban expansion and its simulation using CA-ANN model in Ulaanbaatar, Mongolia. In *Geocarto International*, pp. 1–16. DOI: 10.1080/10106049.2020.1723714.

Ghosh, Bhaskar; Dutta, Indira Kalyan; Totaro, Michael; Bayoumi, Magdy (2020): A Survey on the Progression and Performance of Generative Adversarial Networks, pp. 1–8. DOI: 10.1109/ICCCNT49239.2020.9225510.

Gonog, Liang; Zhou, Yimin (2019): A Review: Generative Adversarial Networks, pp. 505–510. DOI: 10.1109/ICIEA.2019.8833686.

Gunning, David; Stefik, Mark; Choi, Jaesik; Miller, Timothy; Stumpf, Simone; Yang, Guang-Zhong (2019): XAI-Explainable artificial intelligence. In *Science robotics* 4 (37). DOI: 10.1126/scirobotics.aay7120.

Mansour, Shawky; Al-Belushi, Mohammed; Al-Awadhi, Talal (2020): Monitoring land use and land cover changes in the mountainous cities of Oman using GIS and CA-Markov modelling techniques. In *Land use policy* 91, p. 104414. DOI: 10.1016/j.landusepol.2019.104414.

Peng, Kaifeng; Jiang, Weiguo; Deng, Yue; Liu, Yinghui; Wu, Zhifeng; Chen, Zheng (2020): Simulating wetland changes under different scenarios based on integrating the random forest and CLUE-S models: A case study of Wuhan Urban Agglomeration. In *Ecological Indicators* 117, p. 106671. DOI: 10.1016/j.ecolind.2020.106671.

Pontius, Robert Gilmore; Boersma, Wideke; Castella, Jean-Christophe; Clarke, Keith; Nijs, Ton de; Dietzel, Charles et al. (2008): Comparing the input, output, and validation maps for several models of land change. In *Ann Reg Sci* 42 (1), pp. 11–37. DOI: 10.1007/s00168-007-0138-2.

Roodposhti, Majid Shadman; Hewitt, Richard J.; Bryan, Brett A. (2020): Towards automatic calibration of neighbourhood influence in cellular automata land-use models. In *Computers, Environment and Urban Systems* 79, p. 101416. DOI: 10.1016/j.compenvurbsys.2019.101416.

Shafizadeh-Moghadam, Hossein; Asghari, Ali; Tayyebi, Amin; Taleai, Mohammad (2017): Coupling machine learning, tree-based and statistical models with cellular automata to simulate urban growth. In *Computers, Environment and Urban Systems* 64, pp. 297–308. DOI: 10.1016/j.compenvurbsys.2017.04.002.

Uuemaa, Evelin; Antrop, Marc; Roosaare, Jüri; Marja, Riho; Mander, Ülo (2009): Landscape metrics and indices : an overview of their use in landscape research. In *LIVING REVIEWS IN LANDSCAPE RESEARCH* 3 (1), pp. 1–28. Available online at http://biblio.ugent.be/publication/695518.

Wang, Haijun; Guo, Jiaqi; Zhang, Bin; Zeng, Haoran (2021a): Simulating urban land growth by incorporating historical information into a cellular automata model. In *Landscape and Urban Planning* 214, p. 104168. DOI: 10.1016/j.landurbplan.2021.104168.

Wang, Jinzhu; Hadjikakou, Michalis; Bryan, Brett A. (2021b): Consistent, accurate, high resolution, long time-series mapping of built-up land in the North China Plain. In *GIScience & Remote Sensing*, pp. 1–17. DOI: 10.1080/15481603.2021.1948275.

Wang, Rong; Feng, Yongjiu; Wei, Yongliang; Tong, Xiaohua; Zhai, Shuting; Zhou, Yilun; Wu, Peiqi (2021c): A comparison of proximity and accessibility drivers in simulating dynamic urban growth. In *Transactions in GIS* 25 (2), pp. 923–947. DOI: 10.1111/tgis.12707.

Wang, Zhengwei; Healy, Graham; Smeaton, Alan F.; Ward, Tomás E. (2020): Use of Neural Signals to Evaluate the Quality of Generative Adversarial Network Performance in Facial Image Generation. In *Cogn Comput* 12 (1), pp. 13–24. DOI: 10.1007/s12559-019-09670-y.

Xu, Qiantong; Huang, Gao; Yuan, Yang; Guo, Chuan; Sun, Yu; Wu, Felix; Weinberger, Kilian (2018): An empirical study on evaluation metrics of generative adversarial networks. DOI: 10.48550/arXiv.1806.07755.