Response to the Anonymous Referee #3

General comment:

The authors used the STSR-Seg method to develop a novel BRA dataset covering 2016-2021 with a temporal interval one-year, it has been demonstrated to have better performance than Zhang et al. (2022) results and covers the whole China (rural and urban areas), with an overall accuracy of 82.85%. The manuscript has been well-written and clearly organized. However, I still have several concerns about the current manuscript as follows

The authors describe an effort to create building footprint data for all of China. Their dataset is a raster dataset at 2.5m resolution, derived from 10m Sentinel-2 data. They use a deep learning approach involving super-resolution segmentation, allowing to downscale the information from 10m to 2.5m resolution.

The contribution is timely and very relevant, as it tackles several gaps in the global data landscape on human settlements and built-up areas: 1) The created data covers China (including its rural areas), unlike other data products; 2) The dataset is multitemporal (2016-2021) which is rare, allowing for the assessment of built-up growth and shrinkage due to demolition etc.

The paper is well-written and structured. It is very detailed and includes a thorough accuracy assessment against other datasets including a comparison to global datasets available at lower spatial resolution, including datasets from different sources, and also involves hand-crafted validation data and manual checks. The obtained accuracy estimates are quite high and promising.

As I cannot judge the quality of the deep learning framework, I have mostly minor comments, as well as some comments on the data themselves and a request for clarification on the accuracy assessment.

Response:

We are very grateful for the reviewers' comments, which have been very helpful and important in improving the quality of our work. To improve readability, we respond to reviewers' comments in three sections below, i.e., “Part 1: Comments on the data”, “Part 2: Accuracy assessment, comparison” and “Part 3: Minor comments”.

The reviewer's comments are presented in black font, while our responses are presented in blue font. The revised manuscript is presented in red color, and our modifications are highlighted in yellow.

Part 1: Comments on the data

Comments on the data 1:

Empty raster datasets such as CBRA_2016_E113.5_N51.3.tif or CBRA_2016_E76.0_N33.8.tif should be excluded from the dataset.

Response:

Thank you for the reviewer’s comment. The empty raster are regions without buildings, they are now excluded in the final version of the CBRA.

Comments on the data 2:

Until looking at the data, I was very positive towards this manuscript. However, I then had a look at a small selection of the data, and was quite surprised to see very coarse "blobs" delineating settlement areas, rather than mapping "rooftop areas" as the dataset suggests (example in the figure below). I zoomed into 3-4 regions, and most seem to be finer-grained than these blobs and actually delineating individual buildings / rooftops. However, the authors should be transparent and also show such an example in their manuscript, to highlight that the method does not seem to work well everywhere – and possibly provide an explanation for this. Looking at this specific example, I don't think it is defendable to call this "rooftop areas" – this is a quite generalized settlement area, slightly more refined than the GHS-BUILT 10m dataset, shown for comparison.

Response:

Thank you for the reviewer's comment and careful inspection.

Our method employs convolutional neural networks (CNNs) and utilizes super resolution (SR) and semantic segmentation (SS) techniques, which enable us to generate 2.5 m building rooftop results using only 10 m Sentinel-2 images. The following is a simplified diagram of the model architecture (including only the forward inference process).

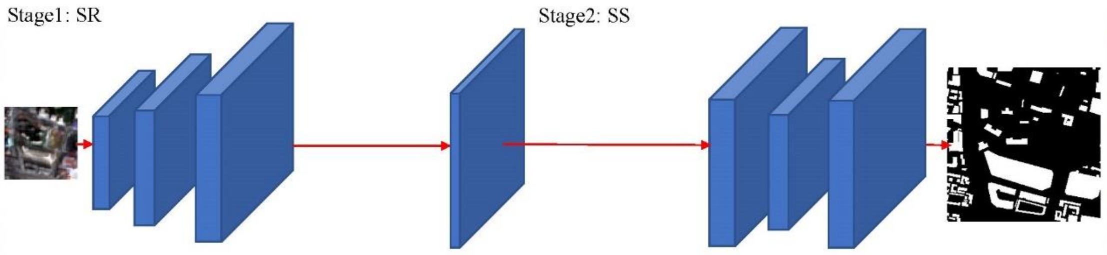


Figure C1: A brief structure of the proposed STSR-Seg, mainly including two parts, i.e., super resolution (SR) and semantic segmentation (SS). In our manuscript, the SR is EDSR module and the SS is the U-net module.

Since the spatial resolution of our results is 2.5 meters, it is difficult to distinguish single buildings when the distance between two buildings is less than 2.5 meters. Furthermore, the deep-learning-based approach we used naturally introduces local ambiguity at the edges of the building, resulting in a buffer-like effect of a few pixels at the building edges (Ding et al., 2021; Zorzi et al., 2021; Guo et al., 2022; Liu et al., 2022). We have included examples of this effect in Fig. C2, where we present an example of boundary localization results using the Unet++ method (Zhou et al., 2019). Even in very high-resolution images (0.3 m), the building boundary extraction results exhibit an uncertain offset of several pixels.

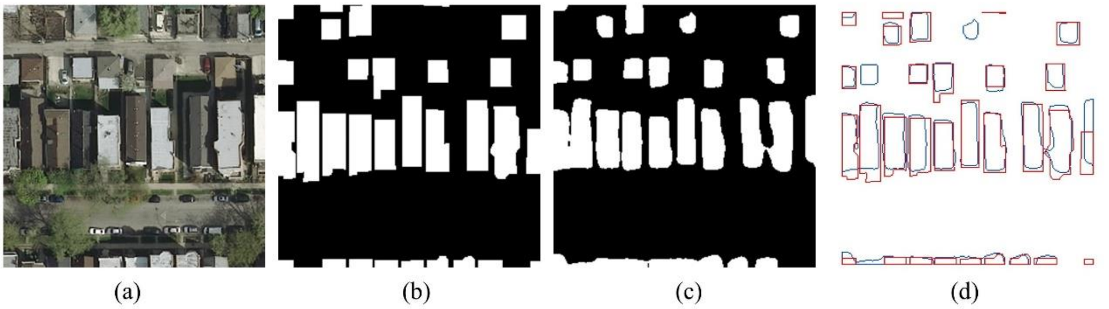


Figure C2: An example of the ambiguity of building edges from Inria building dataset (0.3 m) by the Unet++ method (Zhou et al., 2019). (a) The input image (0.3 m). (b) The ground truth. (c) The extraction result. (d) The boundary result, red is ground truth and blue is prediction. We can see even in the very high-resolution images (0.3 m), the building boundary extraction results still shows with an offset of several pixels.

In densely residential areas, the aforementioned ambiguity is further compounded, as demonstrated in Fig. C3. When buildings stand closely (e.g., less than 6-7 m, i.e., 2-3 pixels at 2.5 m resolution), our method may not be able to effectively distinguish between them. It is a common phenomenon in building extraction task, but the previous work mainly uses very high-resolution images (less than 1 m), making their results seem better than ours.

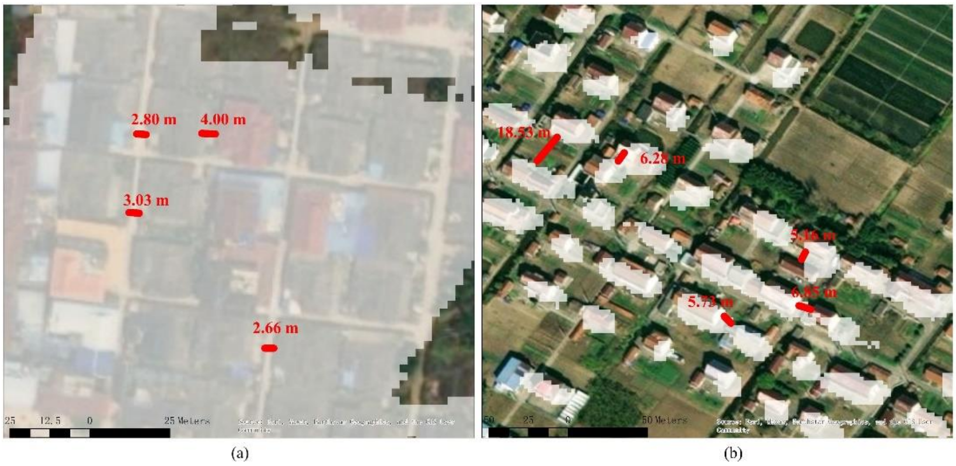


Figure C3: Two examples of blob-like areas and the measured distances. (a) Densely residential area (101.302089ºE, 21.298532ºN). (b) Relatively discrete residential area (121.634662 ºE, 31.746674 ºN). Imagery © 2023 Maxar Techonlogies.

Regarding the potential issue of the inherent ambiguity in edge extraction, we ever conducted an offline experiment to explore other methods that could address this issue, such as RNN-based methods, GAN-based methods, and post-processing methods. We kept the super-resolution module constant and only replaced these methods with our semantic classification head (i.e., the SS part in Fig. C1). However, the results were not satisfactory.

For RNN-based methods, we utilized the Polygon-RNN (Castrejon et al., 2017). However, we found the model got a trick solution whatever the input, and could not be trained successfully (Fig. C4). For GAN-based methods and post-processing methods, we tried the recently introduced ASLNet (Ding et al., 2021) and FrameField (Girard et al., 2021), and the results were also not good compared with our method. The ASLNet could only obtained 27.64% IoU and the FrameField could only obtain 16.67%, while our method reported in the manuscript is 45.51%. We believe these kinds of methods are all designed for high-resolution building extractions (e.g., less than 1 m), thus showing inferior performance in Sentinel-2 imagery.

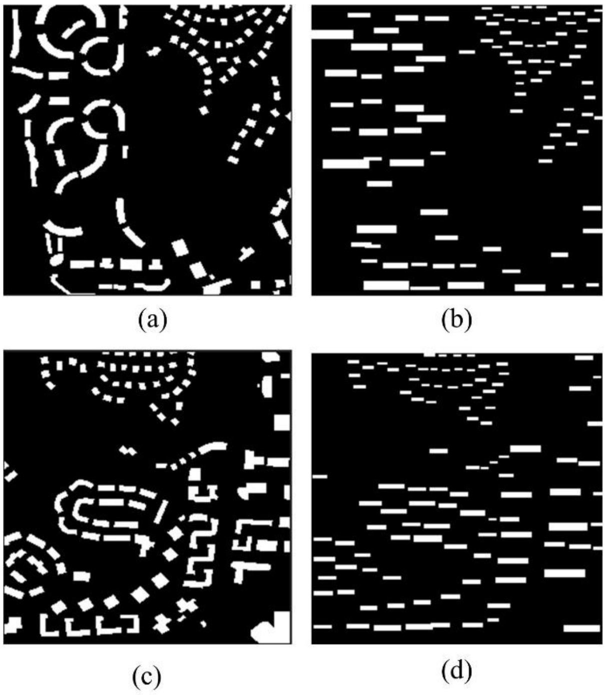


Figure C4: The result of Polygon-RNN, we replace the SS part in Fig. C1 with Polygon-RNN. (a) and (c) are the ground truth, (b) and (d) are the model prediction results. It could be seen that the method learns a trick solution to the input.

Finally, we present the possibility map of our method output in Fig. C5. In our manuscript, we used a threshold of 0.5 to binarize this map, following common practice. However, using a higher threshold in areas with dense building clusters may improve the method's performance in these challenging scenarios. We are now working to provide CBRA's building rooftop probability product (to supplement the published data with the original data link) later to serve the need for more precision.

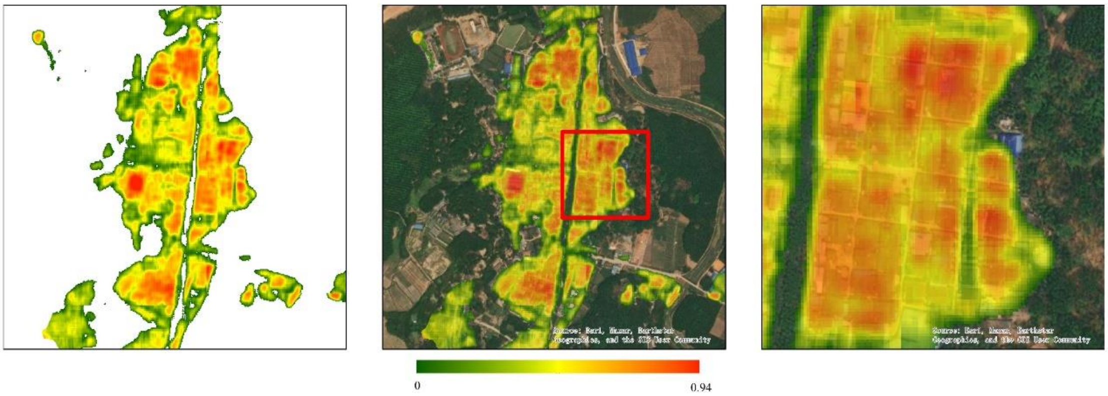

Source: Esri, Maxar, Earthstar Geographics, and the GIS User Community
Source: Esri, Maxar, Earthstar Geographics, and the GIS User Community
0
0.94

Figure C5: (a) The possibility map of the failure area (101.302089ºE, 21.298532ºN). (b) The possibility map with high-resolution imagery from ArcGIS online. (c) An enlargement of the possibility map. It could be seen that the building rooftop area has a higher possibility (red colour). In our study, we used a threshold of 0.5 to obtain the ultimate binary building rooftop extraction outcome, which is consistent with the standard practice of building extraction. However, increasing the threshold value could potentially improve the results for identifying highly dense buildings in rural regions of China. Imagery © 2023 Maxar Techonlogies.

To summarize, the ambiguity inherent in edge extraction results in a blob-like shape in our proposed method, with a potential offset of 1-3 pixels. Despite trying other methods specifically designed for accurate building boundary delineation, we found that they were not successful in Sentinel-2 imagery. It should be noted that our method may not be suitable for identifying very dense buildings in rural areas of China, but we are working to provide the probability map in the near future (we will be releasing a new dataset of CBRA probability maps, which is still in production. We will add a link to the probability map product on the CBRA dataset page when it is ready). Despite this limitation, our method is the first to use Sentinel-2 10 m imagery to achieve long-term and large-scale building rooftop mapping at 2.5 m resolution. This low-cost and dynamic building mapping strategy has not been previously achieved.

We now strengthen the statement of limitation (Sect. 6.2) in the revised manuscript (Fig. 19 in the revised manuscript is the Fig. C3 above), from line 608 to 614 on page 30:

Although our STSR-Seg framework is scalable, allowing larger areas to be monitored (e.g., national scale), there remain some limitations in our approach. Specifically, the segmentation results for densely populated residential areas may present certain rooftops as a single block, rather than individual buildings. Our analysis suggests that this occurrence is primarily due to the consequence of the limited spatial resolution of the results, i.e., 2.5 m. Furthermore, the semantic segmentation technique utilized in the approach may introduce some uncertainty at the edges of buildings, resulting in additional pixels, up to three pixels, at the boundary. Consequently, up to 7.5 meters of buffering may occur, exacerbating the problem of building adhesion. Examples of this issue are presented in Fig. 19…….

References:

Ding, L., Tang, H., Liu, Y., Shi, Y., Zhu, X. X., & Bruzzone, L. (2021). Adversarial shape learning for building extraction in VHR remote sensing images. IEEE Transactions on Image Processing, 31, 678–690.

Zorzi, S., Bittner, K., & Fraundorfer, F. (2021). Machine-learned regularization and polygonization of building segmentation masks. 2020 25th International Conference on Pattern Recognition (ICPR), 3098–3105.

Guo, H., Du, B., Zhang, L., & Su, X. (2022). A coarse-to-fine boundary refinement network for building footprint extraction from remote sensing imagery. ISPRS Journal of Photogrammetry and Remote Sensing, 183, 240–252.

Liu, Z., Tang, H., & Huang, W. (2022). Building Outline Delineation From VHR Remote Sensing Images Using the Convolutional Recurrent Neural Network Embedded With Line Segment Information. IEEE Transactions on Geoscience and Remote Sensing, 60, 1–13.

Zhou, Z., Siddiquee, M. M. R., Tajbakhsh, N., & Liang, J. (2019). Unet++: Redesigning skip connections to exploit multiscale features in image segmentation. IEEE Transactions on Medical Imaging, 39(6), 1856–1867.

Castrejon, L., Kundu, K., Urtasun, R., & Fidler, S. (2017). Annotating object instances with a polygon-rnn. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 5230–5238.

Girard, N., Smirnov, D., Solomon, J., & Tarabalka, Y. (2021). Polygonal building extraction by frame field learning. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 5891–5900.

Comments on the data 3:

I would anticipate much wider usage of the data if they were provided as vector data (i.e. polygon objects describing each building) rather than raster data. The fact that the authors provide 2.5m-raster data still leave a major chunk of processing work to the user. While there are applications where the fine-grained raster data is useful, most applications will be based on vector data. The authors correctly mention the vectorization step as “future work”, but I would like to raise the discussion here if it would be beneficial to do this at this point, or otherwise provide a vectorized version of the data in the near future – just as some “food for thought”.

Responses:

Thank you for your valuable comment.

We acknowledge the significance of the vector results in building rooftop extraction. However, we have identified limitations in our results, particularly in the segmentation of single buildings in densely populated residential areas, as highlighted in our response to "Comments on the data 2", which is why we decided not to publish the vector version of the CBRA until we can overcome these limitations and ensure the highest quality of data for users.

However, we understand the significance of vector data for many applications, and we plan to address this in the near future by performing the following tasks:

(1) We will provide the vector version of the CBRA using commonly used vectoring algorithms. This initiative is expected to facilitate more extensive utilization of the data and ease its incorporation into research projects (we will be releasing a new vector version for CBRA, which is still in production. We will add a link to this product on the CBRA dataset page when it is ready).

(2) We will make our source code and training data publicly available after the completion of the review process. This will enable researchers to replicate and build upon our work and encourage the development of new methods that can achieve higher accuracy in building rooftop extraction with relatively coarse resolution imagery (the code will be found in https://github.com/zpl99/STSR-Seg, we have already mentioned in the Introduction of the manuscript).

(3) We plan to develop new methods that can achieve higher accuracy in building rooftop extraction, particularly in densely residential areas.

We now rewrite the Sect. 6.2 “Limitations and prospects” to be more specific, from 615 to 639 on page 30:

……Besides, there is a need for further improvement in the delineation of the building boundaries within the CBRA. Buildings differ from other objects of interest in that they have regularized boundaries (e.g., polygons made of lines and vertexes). However, our dense pixel-to-pixel classification method disregards the morphology of the building, resulting a blob-like shape. For example, in Table 5……This indicates that the CBRA results suffer from ambiguous localization on the building boundaries.

We have noticed that there are many studies on the morphology extraction of buildings in recent years, such as instance segmentation methods (Liu et al., 2022; Zhu et al., 2021; Huang et al., 2021a). We also try to replace our semantic segmentation branch with current instance segmentation methods, e.g., recurrent neural network methods (Liu et al., 2022). However, the results are not good and even fail in our off-line experiment, mainly because these methods are designed for very high-resolution aerial images (sub-metric level). In addition, the efficiency of these methods is too low to support national-level building mapping.

……Many endeavors utilize a post-processing strategy, e.g., Douglas–Peucker algorithm, to achieve regularization (Wei et al., 2019; Chen et al., 2020; Zorzi et al., 2021) and such strategy has shown the success in building mapping at a relatively small scale (Wei et al., 2019). However, in the CBRA, the use of post-processing would introduce errors due to several block estimations in the densely residential area as aforementioned. Considering the potential errors by vectorization, it is hard to provide vector results of the CBRA.

The CBRA provides full-coverage and multi-annual information of building rooftops for China at 2.5 m spatial resolution, and the proposed STSR-Seg offers an opportunity to obtain high-resolution output by using relative low-resolution remote sensing images. However, our findings are constrained by the adhesion of closely located buildings and the blob-like shapes of rooftops. In the near future, we aim to enhance our methodology by designing more powerful model architecture and utilizing multisource data, including synthetic-aperture radar (SAR), and other BRA datasets, with the goal of achieving vector outputs.

Part 2: Accuracy assessment, comparison

Accuracy assessment, comparison # 1:

Table 4: Why is there only recall reported for the rural scenes, whereas for the urban scenes you report recall, F1, Iou, OA? And why you do not report Precision in both cases? This need to be done and is standard for an accuracy assessment. Of course the reader could calculate the precision based on recall and F1, but please provide Precision, recall, OA, F1, IoU for both the rural and the urban scenario. No rationale is provided for only reporting recall in the rural scenario.

Response:

Thank you for the reviewer's comment.

Regarding our decision to report only recall in the rural scenario, we wish to clarify that this was primarily due to limitations in our test sample selection. Specifically, while we were able to utilize accurate and reliable test samples in urban areas by employing vector data from the National Platform for Common Geospatial Information Service of China, we encountered difficulties in identifying equally reliable validation samples for rural areas.

In order to address this issue, we turned to building distribution data sourced from Open Street Map (OSM) (line 217), which constitutes a form of volunteer geographic information data. However, given the imperfect development of such data in the Chinese region, we recognized that it is subject to errors. To mitigate this, we corrected the OSM data based on the "World Imagery" provided by ArcGIS online, although it should be noted that this imagery does not provide a specific acquisition time (line 218).

Despite these limitations, we made every effort to ensure the accuracy of our existing test samples, and we therefore opted to report recall as the primary metric in rural areas. This was because we could accurately calculate the true positives (TP) and false negatives (FN) based on the existing test samples in the rural area. It should be noted, however, that the calculation of false positives (FP) and true negatives (TN) may be subject to some bias, as certain background pixels which should be classified as building pixels may not have been successfully identified by our visual inspection due to the uncertainty of acquisition time of high-resolution images.

We now add more explanations in the revised manuscript.

In "Section 3.2 Reference data", from 219 to 220 on page 10:

……by ArcGIS Online (Arcgis online, 2022). Despite our efforts, the accuracy of our interpretation is subject to some omissions due to the uncertainty of acquisition time of the images used. Finally, building rooftops of 14 villages are obtained (30, 000 buildings), as shown in Fig.

4…….

For the caption of Table 4 on page 19:

Table 4: Performance metrics for building rooftop extraction. Only recall with respect to OSM data is reported in rural areas, due to the challenges of accurately calculating other metrics caused by omissions in the OSM data.

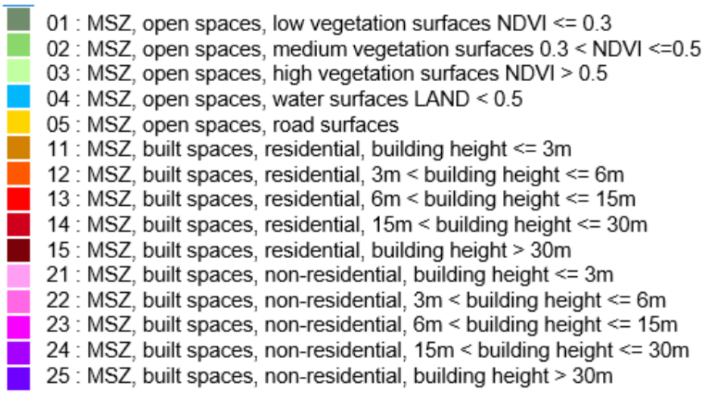
Accuracy assessment, comparison # 2:

Moreover, it is unclear how the accuracy estimates for the Global Human Settlement Layer (GHSL) as reported in Fig. 4 were produced. Which of 10m GHS-BUILT data products was used? There is either the GHS_BUILT_S_E2018_GLOBE_R2022A_54009_10_V1_0 dataset, or the GHS-BUILT-S2 dataset. Both are continuous, with the former reporting the 10m built-up fraction, and the latter reporting the built-up probability. Please provide the following information: Which of the datasets was used? And how were these continuous data thresholded in order to carry out a binary (2-class) agreement assessment? I.e., what cut-off value was used, and how was this cut-off value derived?

Response:

Thank you for the comment.

We used it considering the problem that the threshold is not well delineated. So, the datasets we used is GHS_BUILT_C_MSZ_E2018_GLOBE_R2022A_54009_10_V1_0 product, which provides the category labels for each pixel, as shown in the figure below. For the specific processing, we reclassify this raster data, i.e., 01-05 is assigned to 0 and 11-25 is assigned to 255.



🟩 01 : MSZ, open spaces, low vegetation surfaces NDVI <= 0.3
🟩 02 : MSZ, open spaces, medium vegetation surfaces 0.3 < NDVI <=0.5
🟩 03 : MSZ, open spaces, high vegetation surfaces NDVI > 0.5
🟦 04 : MSZ, open spaces, water surfaces LAND < 0.5
🟨 05 : MSZ, open spaces, road surfaces
🟧 11 : MSZ, built spaces, residential, building height <= 3m
🟧 12 : MSZ, built spaces, residential, 3m < building height <= 6m
🟥 13 : MSZ, built spaces, residential, 6m < building height <= 15m
🟥 14 : MSZ, built spaces, residential, 15m < building height <= 30m
🟫 15 : MSZ, built spaces, residential, building height > 30m
🟪 21 : MSZ, built spaces, non-residential, building height <= 3m
🟪 22 : MSZ, built spaces, non-residential, 3m < building height <= 6m
🟪 23 : MSZ, built spaces, non-residential, 6m < building height <= 15m
🟪 24 : MSZ, built spaces, non-residential, 15m < building height <= 30m
🟪 25 : MSZ, built spaces, non-residential, building height > 30m

Figure C6: Morphological Settlement Zone (MSZ) Legend (Schiavina et al., 2022)

References:

Schiavina, M., Melchiorri, M., Pesaresi, M., Politis, P., Freire, S., Maffenini, L., Florio, P., Ehrlich, D., Goch, K., & Tommasi, P. (2022). GHSL Data Package 2022.

Accuracy assessment, comparison # 3:

The observed accuracy drop from urban towards rural is typical for settlement mapping, please place your work in the context of the literature, e.g. by citing Leyk et al. 2018, or Kaim et al. 2022.

Response:

Thank you for the reviewer's comment. The literature work is now added to the revised manuscript, from line 474 to 480 on page 20:

……Compared to other datasets that provide information related to buildings in rural areas, the CBRA is at a significantly fine-grained scale, albeit with a greater presence of block areas in rural versus urban environments (Fig. 13).

……Besides, the CBRA has a full coverage of China, including the rural areas at a finer scale than other existing full-coverage and thematic-related products. However, a decline in accuracy in rural areas, consistent with prior studies (Leyk et al., 2018; Kaim et al., 2022), has been observed…….

References:

Leyk, S., Uhl, J. H., Balk, D., & Jones, B. (2018). Assessing the accuracy of multi-temporal built-up land layers across rural-urban trajectories in the United States. Remote Sensing of Environment, 204, 898–917.

Kaim, D., Ziółkowska, E., Grădinaru, S. R., & Pazúr, R. (2022). Assessing the suitability of urban-oriented land cover products for mapping rural settlements. International Journal of Geographical Information Science, 36(12), 2412–2426.

Accuracy assessment, comparison # 4:

The authors use the overall accuracy for their accuracy assessment. However, it is well-known that OA yields biased results in the case of imbalanced class distributions (see Shao et al. 2019, Uhl & Leyk 2022) for a recent in-depth study. Such class imbalance is typically the case for built-up vs not built-up assessments, in particular in rural areas. Under the light of this potential bias, please add some sentences critically evaluation the magnitude of the OA values obtained. That being said, I appreciate the authors also report IoU and F-1.

Response:

Thank you for the reviewer's suggestion. We now add more descriptions about this issue, from 441 to 444 on page 18:

……For recall, the CBRA obtains 74.66%, which achieves great improvement (+ 27.29%) compared with 90-cities-BRA, mainly because our robust designation of STSR-Seg framework. It is noteworthy that solely relying on OA for evaluating the performance of CBRA is inadequate due to the category-unbalanced nature of building roof extraction. The OA score may introduce a potential bias in this scenario (Shao et al., 2019; Uhl and Leyk, 2022), and therefore, multiple metrics must be utilized when assessing the performance of CBRA.

References:

Shao, G., Tang, L., and Liao, J.: Overselling overall map accuracy misinforms about research reliability, Landsc Ecol, 34, 2487–2492, 2019.

Uhl, J. H. and Leyk, S.: A scale-sensitive framework for the spatially explicit accuracy assessment of binary built-up surface layers, Remote Sens Environ, 279, 113117, 2022.

Accuracy assessment, comparison # 5:

Fig. 16: Legend should be swapped – the blue should be on the left, and red on the right, also in Fig. 18a.

Response:

Thank you for your careful inspection, the Fig. 16 and 18 now are corrected in the revised manuscript.

Accuracy assessment, comparison # 6:

Fig. 17 b and c: I don't understand what is the difference between panel b) and c), besides the different visualization technique. Please clarify. Moreover, I don't think the pie charts are a good choice here. They don't show the change over time. Please use a layer plot for b) just, as you did in c).

Response:

Thank you for your valuable comment.

The purpose of including subfigure (b) was to illustrate the variation in the proportion of rooftop area of buildings in different urban clusters of China, with respect to the total national rooftop area. We acknowledge that this subfigure may appear redundant with subfigure (c). Moreover, the pie chart representation may not be the most appropriate for displaying changes over time.

In response to your feedback, we have integrated subfigure (b) and (c) into a single figure in the form of a stacked bar chart, which is presented below for your review:

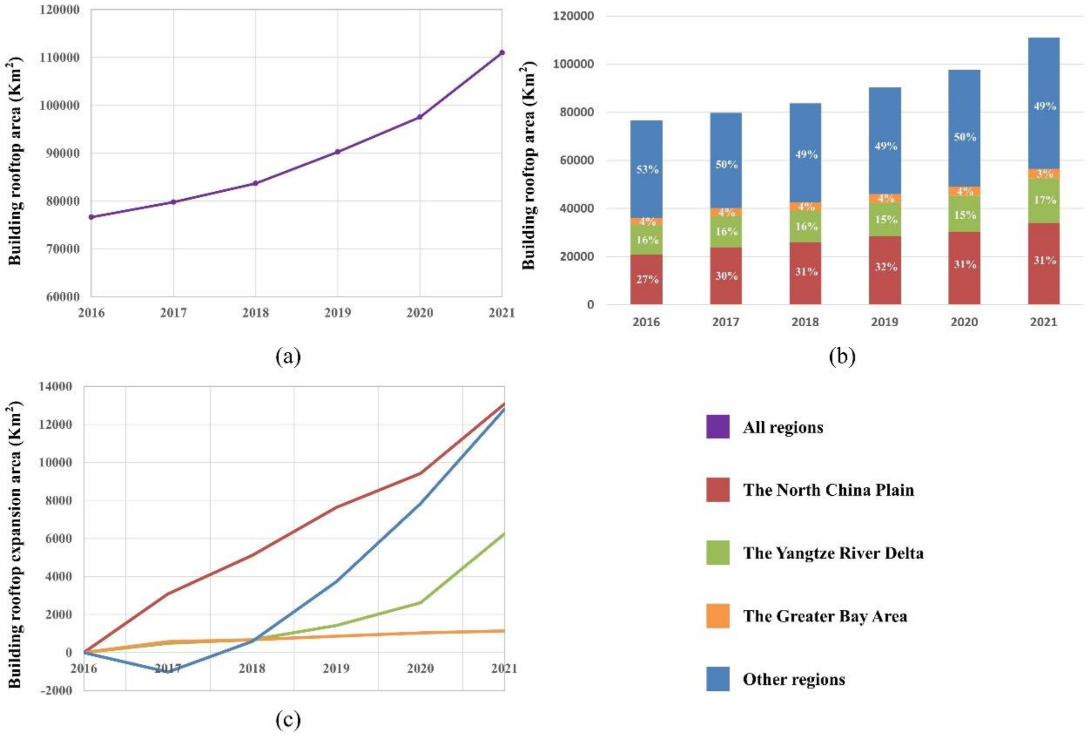


Figure 17: The change of building rooftop area of China and three biggest city clusters in China (NCP, YRD and GBA) over the period of 2016-2021. (a) The annual statistic of building rooftop area in China. (b) The proportion of building rooftop of the biggest city clusters in China and other regions from 2016 to 2021. (c) The increased building rooftop area on each city clusters and other regions.

Accuracy assessment, comparison # 7:

Fig. 18: the green color used to show demolition is different in the map and in the legend.

Response: Thank you for your careful inspection, the 18 now is updated with consistent color in the revised manuscript.

Accuracy assessment, comparison # 8:

Fig. A2, caption: Figure A2: The probability density distribution. …. Of what???

Response: Thank you for your comment. It should be the probability density distribution of the ratio of buildings to built-up area. We compute the probability density distribution of this ratio and observe that it approximates a Gaussian distribution. Leveraging this prior knowledge, we train the model to extend its coverage to areas where high-resolution samples are unavailable. We now add more details to the caption of Fig. A3 in the revised manuscript.

Part 3: Minor comments

Minor comments #1: Please provide a rationale for using the term “rooftop area” instead of “building footprint area” or “built-up area”

Response:

Thank you for the reviewer’s suggestion.

In our research, we aim to extract individual buildings from Sentinel-2 satellite images using super-resolution techniques which help enhance the original resolution of 10 m to a new resolution of 2.5 m. We appreciate the suggestion to consider the use of alternative terms such as "built-up area" or "building footprint area."

The term "built-up area" refers to the total area of land that has been developed or modified by human activity, including buildings, roads, and other infrastructure. However, as our research focuses on the specific features of individual buildings rather than the broader urban environment, we found that the term "built-up area" was not appropriate for our study.

"Building footprint" refers to the total area that a building covers on the ground, including any exterior walls or other structural elements that extend beyond the building's primary enclosed space. While this term is useful for understanding the physical layout of a building, we found it challenging to directly extract building footprints from Sentinel-2 data due to resolution limitations. As an alternative, we chose to use the term "rooftop area," which specifically refers to the top surface of a building and provides a more practical option for our research.

We acknowledge that both "building footprint area" and "rooftop area" are specific to individual buildings and may be appropriate in different contexts. However, in our research, we found that "rooftop area" was a suitable term to describe our focus.

Minor comments #2: Line 75: What means “and F1 score of 2.5 m,” …. I don’t understand what the authors mean here.

Response:

Thank you for the reviewer’s comment. We originally meant to emphasize that the resolution of our results is 2.5 meters, but this sentence is redundant. We now remove the word “2.5 m” in the revised manuscript.

Minor comments #3: Line 106: No need to define an acronym for state-of-the-art; the term is only used twice in the paper.

Response:

Thank you for the reviewer’s suggestion. We remove this acronym in the revised manuscript.

Minor comments #4: Line 159: “apple” ?

Response:

We apologize that this is a spelling error on our manuscript, it should be “apply”. We now correct it in the revised manuscript.

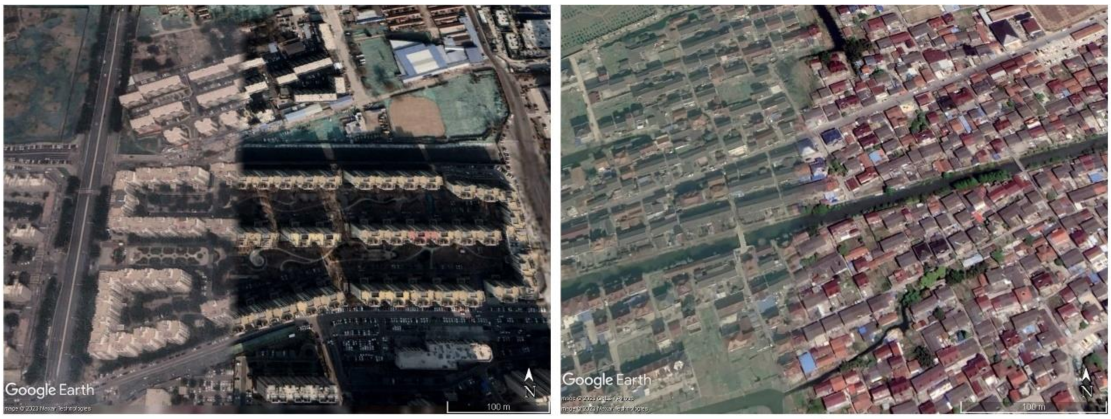
Minor comments #5: Line 161: Please explain what you mean by “geographical offset”.

Response:

Thank you for your inquiry.

The geographical offset we refer to is the bias in the GES imagery. This bias is mainly caused by the following reason:

The acquisition of larger scale GES images generally necessitates the stitching of images from multiple sensors. However, this process can result in large errors at the stitching regions, especially in areas with significant height variations, such as high-rise buildings. Examples of such errors are presented in Fig. C7, which were obtained using Google Earth Pro software.



(a) (b)

Figure C7: Two examples about the stitching part of GES imagery. (a) The high-rise buildings. (b) The low-rise buildings. It could be seen that the offset is more obvious in the high-rise buildings than that in the low-rises. Imagery © 2023 Maxar Techonlogies.

Details regarding this phenomenon are introduced in the revised manuscript (Fig. A1 is the Fig. C7 above), from line 159 to 160 on page 6 :

However, the GES images are collected from various kinds of high-resolution satellites, and have two potential problems when applied to large-scale mapping: (1) inconsistent geographical offset (illustrated in Fig. A1), and (2) inconsistent acquisition time (e.g., the image is obtained from various satellite sensors with different acquisition times) which results in spatio-temporal inconsistency in the generated product.

Minor comments #6: 155-165: nice transition and justification for the contribution of the paper.

Response:

Thank you for your comment.

Minor comments #7: Table 1. Nice overview on existing datasets.

Response:

Thank you for your comment.

Minor comments #8: Fig. 2: Please include the GHS-BUILT dataset here, from the Global Human Settlement Layer, e.g., the GHS-BUILT-10m built up layer. This will provide a nice overview on recent work at a global scale, and highlight the merit of your work.

Response:

Thank you for your comment. The GHS-BUILT-10m built up layer is now added to Fig. 2 on page 6:

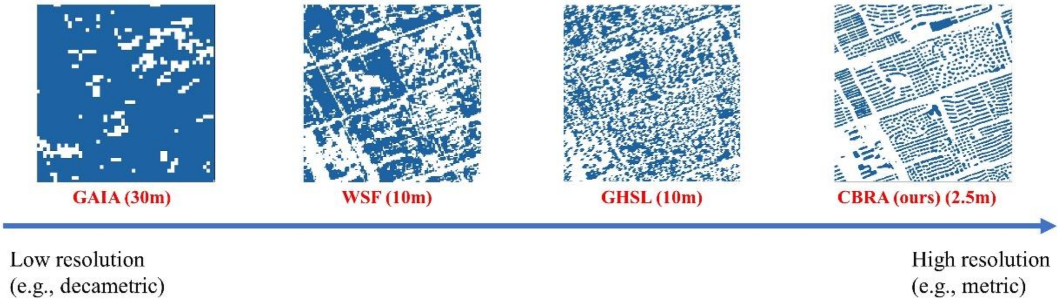


Figure 2: An example of the result from several representative building-related datasets (121.344419°E,31.093870°N). The GAIA (Gong et al., 2020b) reflects the impervious area (30 m). The WSF (Marconcini et al., 2020b) and GHSL (Corbane et al., 2021) are the human settlement data (10 m). The CBRA (ours) is the building rooftop area data (2.5 m).

Minor comments #9: Caption Fig. 2: "cloud under" change to "cloud cover under"

Response:

Thank you for your comment. Correct.

Minor comments #10: Fig. 4: Text is very small, please increase font size, and decrease spacing between lines; this way, the space can be used more efficiently.

Response:

Thank you for your suggestion. Fig. 4 is updated in the revised manuscript, on page 9:

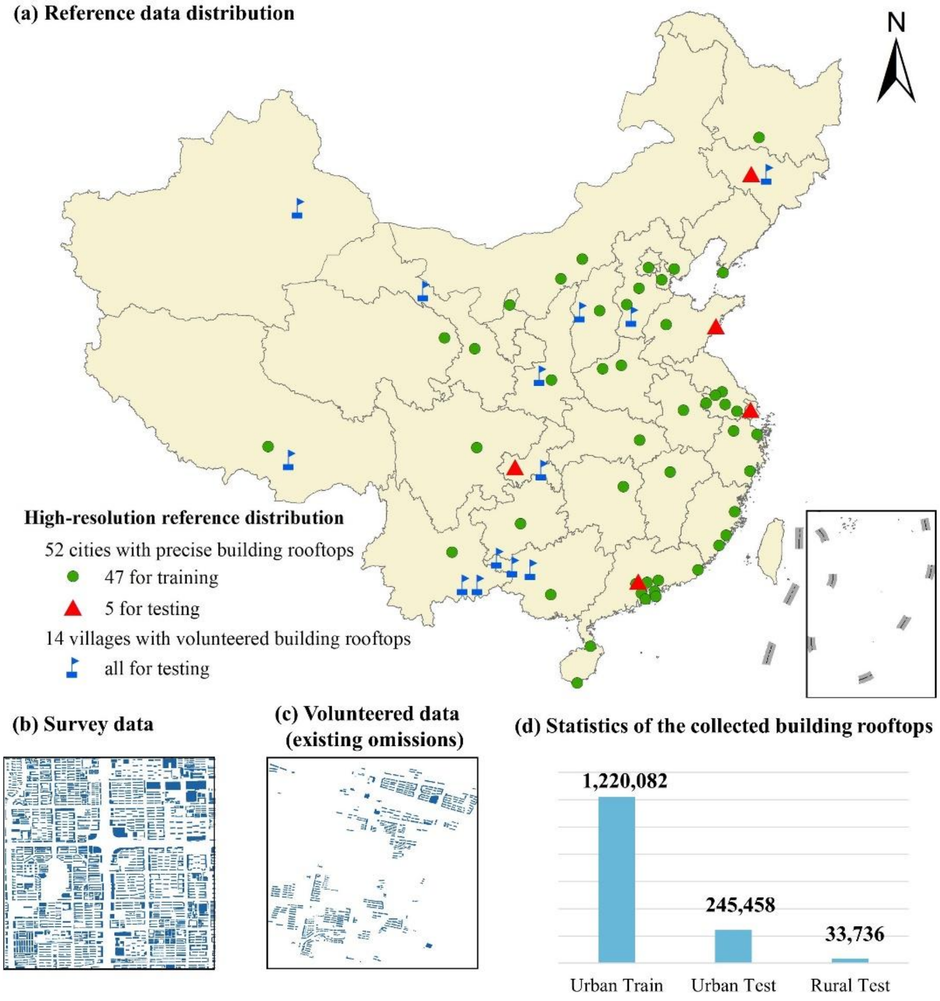

(a) Reference data distribution
N
High-resolution reference distribution
52 cities with precise building rooftops
● 47 for training
▲ 5 for testing
14 villages with volunteered building rooftops
all for testing
(b) Survey data
(c) Volunteered data
(existing omissions)
(d) Statistics of the collected building rooftops
1,220,082
245,458
33,736
Urban Train
Urban Test
Rural Test

Figure 4: Illustration of the collected high-resolution reference. (a) is the high-resolution reference distribution map (base map © OpenStreetMap contributors 2023. Distributed under the Open Data Commons Open Database License (ODbL) v1.0). (b) and (c) are real-world examples of the collected survey data (43.88°N，125.37°E）(survey data © Tiandi-Map) and the volunteered data (33.47°N，119.79°E）(volunteered data © OpenStreetMap contributors 2023. Distributed under the Open Data Commons Open Database License (ODbL) v1.0), respectively. (d) is the statistic of building rooftops.

Minor comments #11: Fig. 9 – caption: "Comparison of the CBRA and the other dataset" – please name the "other dataset".

Response:

Thank you for your comment. We now replace "other dataset" with "90-cities-BRA (Zhang et al., 2022)"

References:

Zhang, Z., Qian, Z., Zhong, T., Chen, M., Zhang, K., Yang, Y., Zhu, R., Zhang, F., Zhang, H., & Zhou, F. (2022). Vectorized rooftop area data for 90 cities in China. Scientific Data, 9(1), 1–12.