Supplement of Earth Syst. Sci. Data, 15, 3547–3572, 2023
https://doi.org/10.5194/essd-15-3547-2023-supplement

Open Access Earth System Science Data

Supplement of

China Building Rooftop Area: the first multi-annual (2016–2021) and high-resolution (2.5 m) building rooftop area dataset in China derived with super-resolution segmentation from Sentinel-2 imagery

Zeping Liu et al.

Correspondence to: Hong Tang (hongtang@bnu.edu.cn)

(a) (b)

Figure S1: Two examples about the stitching part of GES imagery. (a) The high-rise buildings. (b) The low-rise buildings. To acquire larger-scale images, multiple sensors are need to be stitched together, but this process can introduce significant errors at the stitching regions, particularly in areas with height variations like high-rise buildings. The offset is more pronounced in high-rises than in low-rises, as can be seen in the images. Imagery © 2023 Maxar Techonlogies.

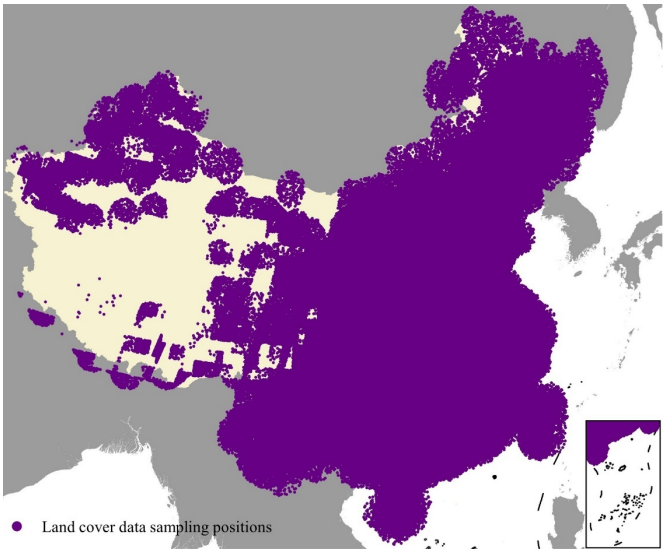


Figure S2: The position of generated land cover samples. The point is sampled from Gaussian distribution with a standard deviation of 150 *km* based on the third level administrative region of China. Base map © OpenStreetMap contributors 2023. Distributed under the Open Data Commons Open Database License (ODbL) v1.0.

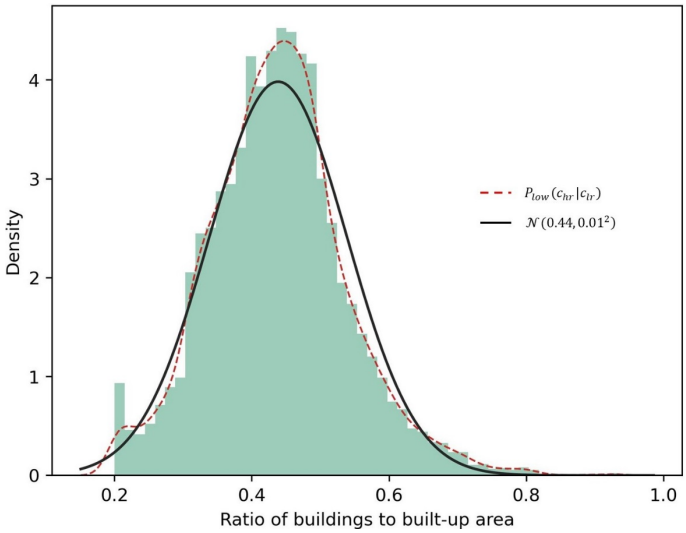


Figure S3: The probability density distribution of the ratio of buildings to built-up area. The red dot curve is our statistical result based on the collected high-resolution survey data and the corresponding low-resolution built land cover data. The black curve represents the estimated Gaussian distribution with 0.44 as μ and 0.01 as σ. It is worth noting that our estimate is only derived from urban areas and is therefore biased. However, as previously shown by (Malkin et al., 2018), the impact of this bias is insignificant because of the robustness of the deep learning model in its modern design.

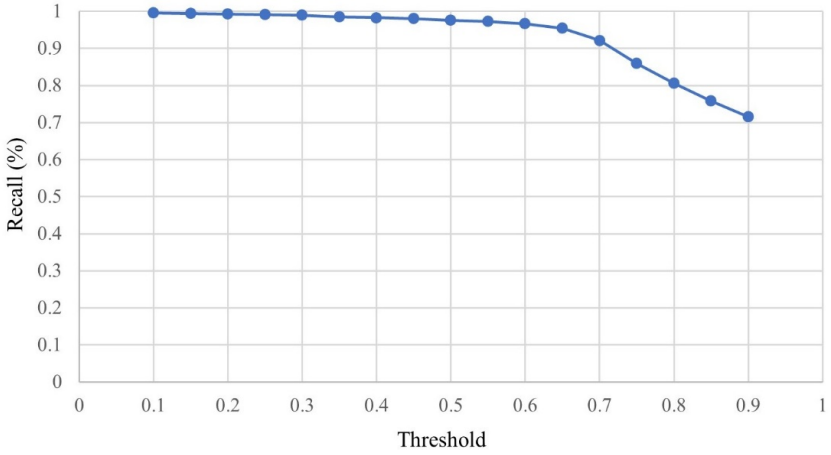
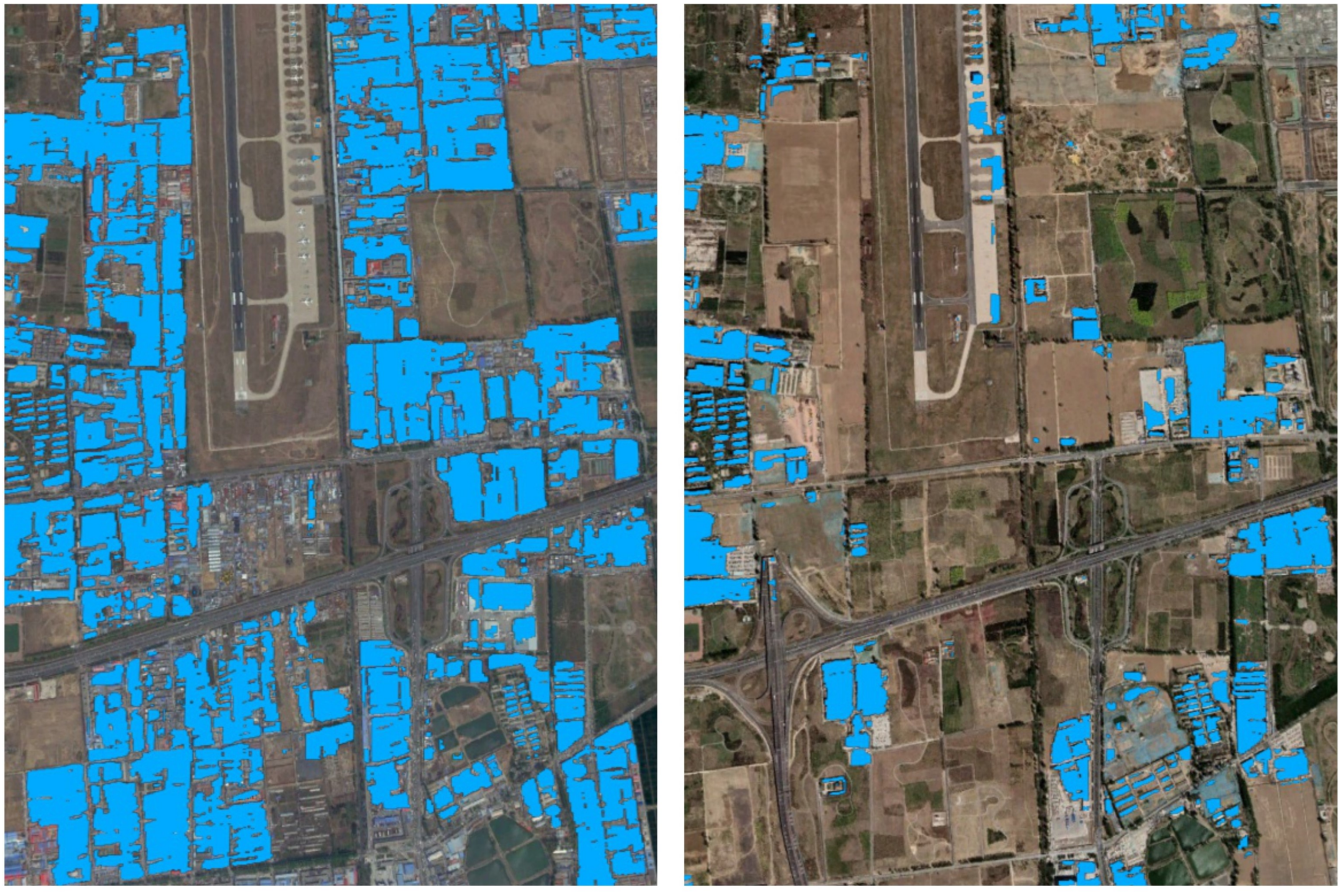
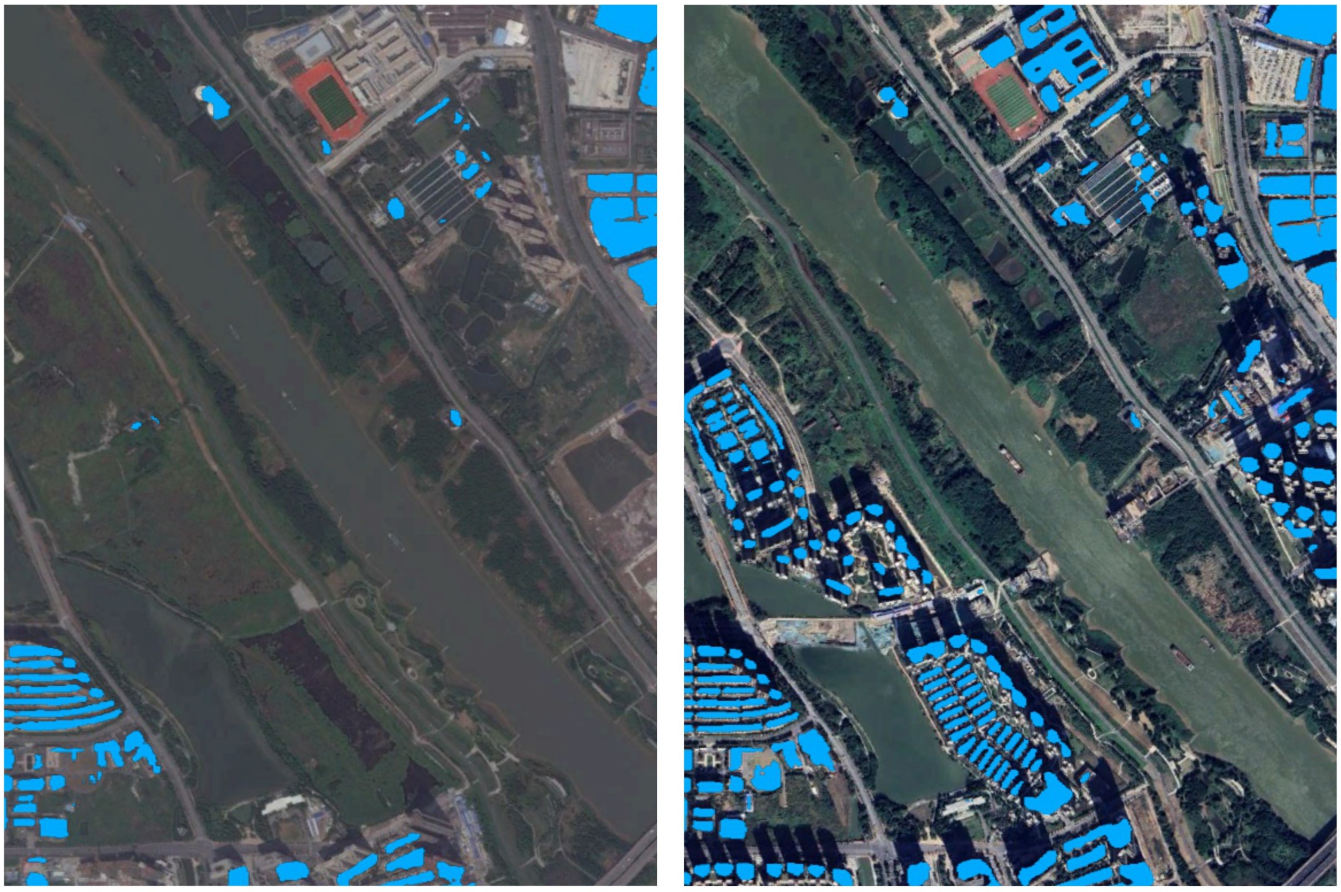


Figure S4: The recall curve for the threshold selection. The threshold interval of 0.05 was used to binarize the Dynamic World product, and it was compared with the building distribution data we collected from 52 cities in China to calculate recall respectively. When the threshold is 0.3, recall reaches 0.99, indicating a threshold value of 0.3 and below ensures that the correct results are not filtered out incorrectly in urban areas. As for the rural area, due to the lack of reliable building rooftop annotations, we judge the thresholds by visual observation of the images, and finally we consider 0.2 is the best option.



(a) (b)

Figure S5: A zoomed result of the demolition process with high-resolution images (116.387674°E, 39.766427°N). (a) The identified building rooftop area of 2016. (b) The identified building rooftop area in 2021. Imagery © 2023 Maxar Techonlogies.



(a) (b)

Figure S6: A zoomed result of the construction process with high-resolution images (113.037558ºE, 23.014343ºN). (a) The identified building rooftop area of 2016. (b) The identified building rooftop area in 2021. Imagery © 2023 Maxar Techonlogies.

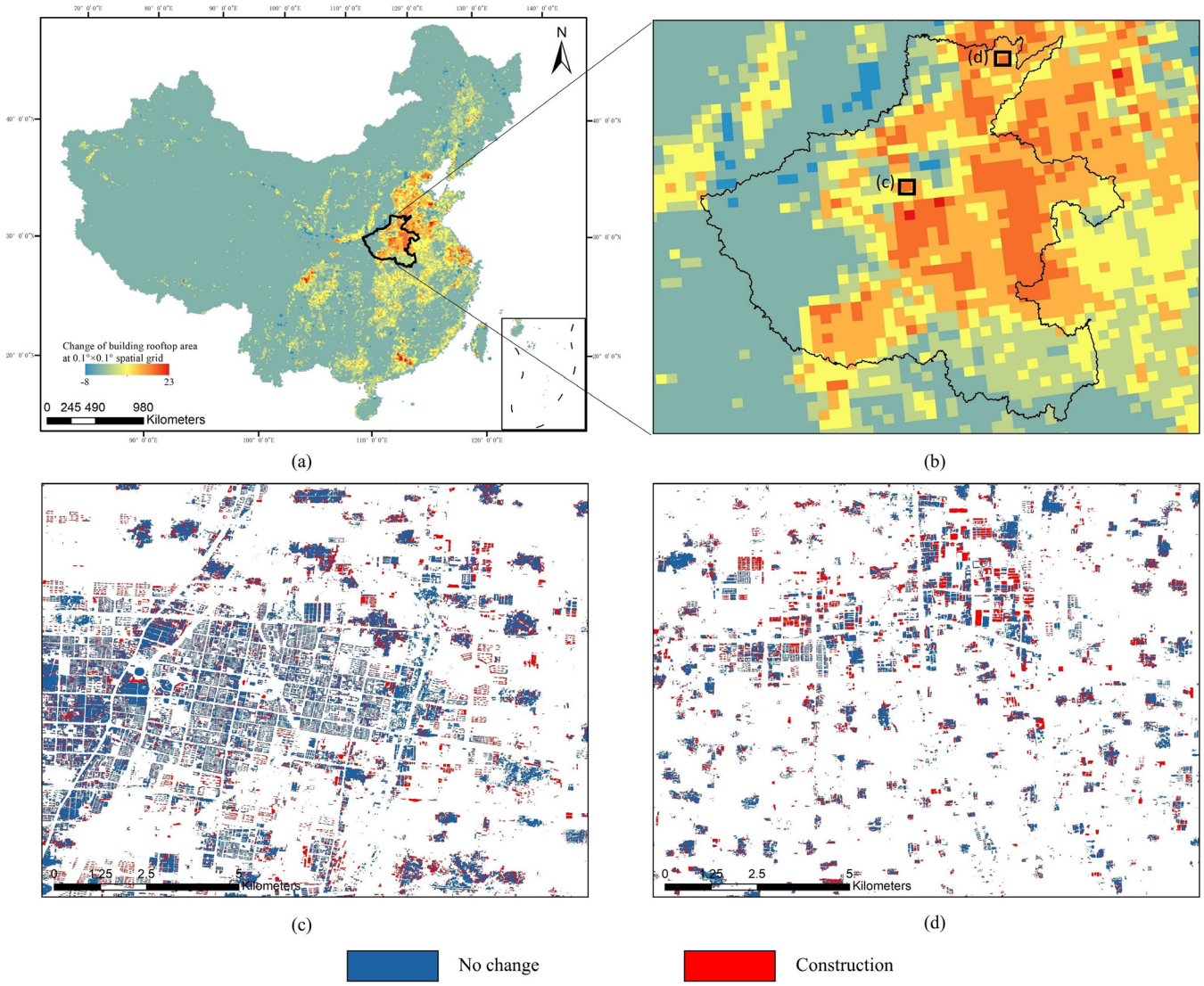


Figure S7: Enlargement of the Henan province of China. (a) The spatial distribution of the annual change of building rooftop area over the period of 2016-2021. The area fraction is aggregated within the $0.10° \times 0.10°$ spatial grid (base map © OpenStreetMap contributors 2023. Distributed under the Open Data Commons Open Database License (ODbL) v1.0). The high-resolution image and our extraction result from 2016. (b) The enlargement of the Henan province. (c) and (d) are two examples of the new construction of buildings from 2016 to 2021, located at "115.087461°E, 35.769057°N" and "113.686257°E, 34.517389°N", respectively.

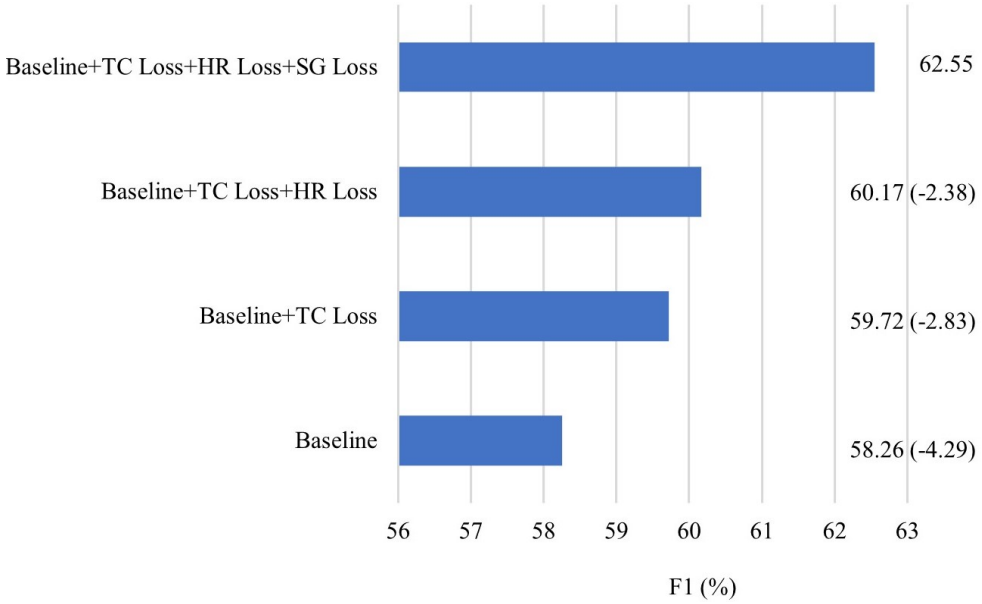


Figure S8: Ablation study of our training method. The first row shows the F1-score performance of our best model including all the learning strategies, i.e., TC loss, HR loss and SG loss. The second row shows the best model without SG loss. The third row shows the best model without HR Loss and SG loss. The last row is our baseline.

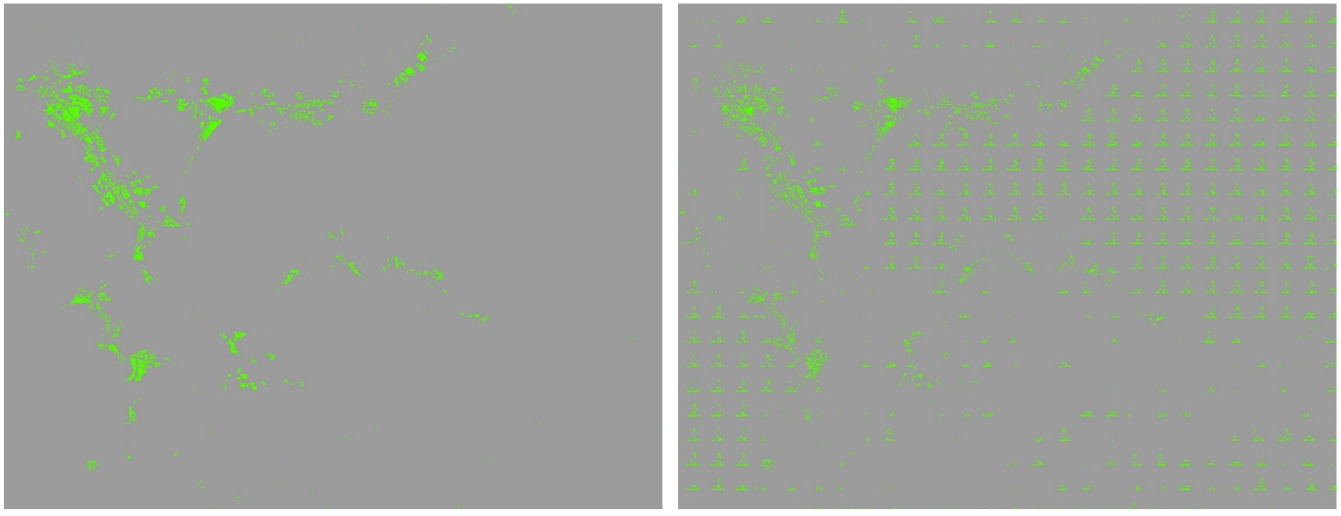


Figure S9: An example of the importance of SG loss. (a) The result of our best model. (b) The result of our best model without SG loss supervision. Without utilizing SG loss, the model only converges on the limited collected survey rooftop data, which limits the generalization performance of the model over a large scale (e.g., national scale), resulting in unexpected false predictions.

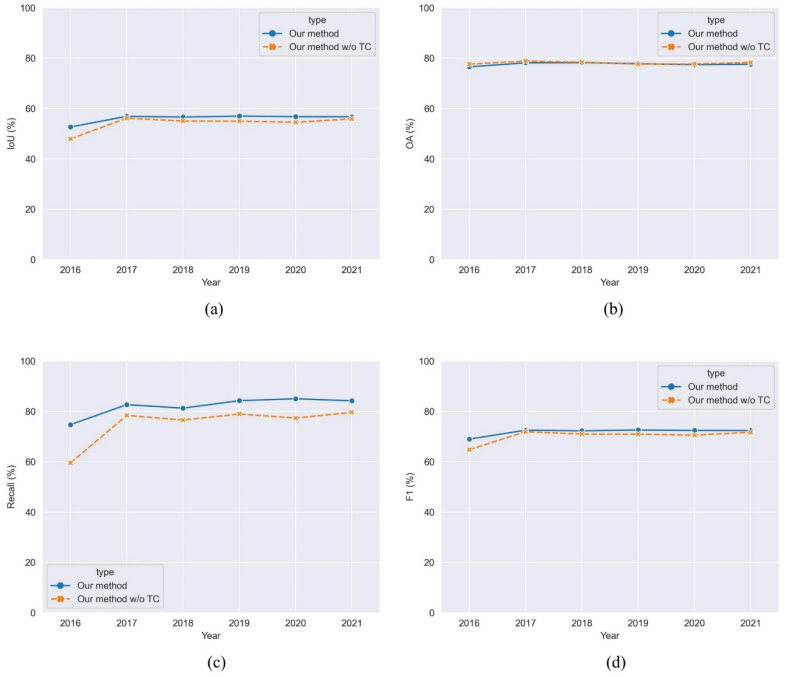


Figure S10: Temporal accuracy in the old town of Beijing of our method and our method without TC loss. (a) IoU. (b) OA. (c) Recall. (d) F1.

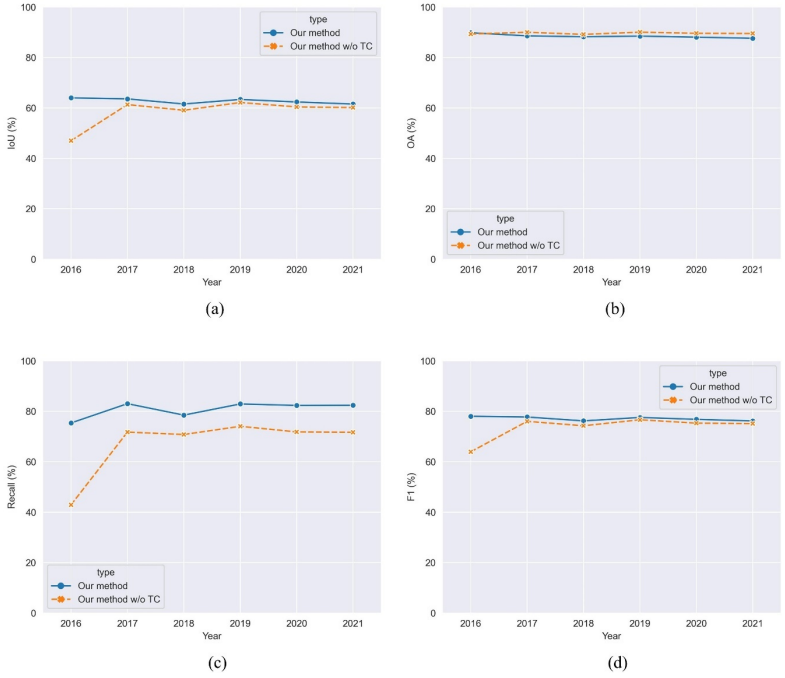


Figure S11: Temporal accuracy in Hong Kong of our method and our method without TC loss. (a) IoU. (b) OA. (c) Recall. (d) F1.

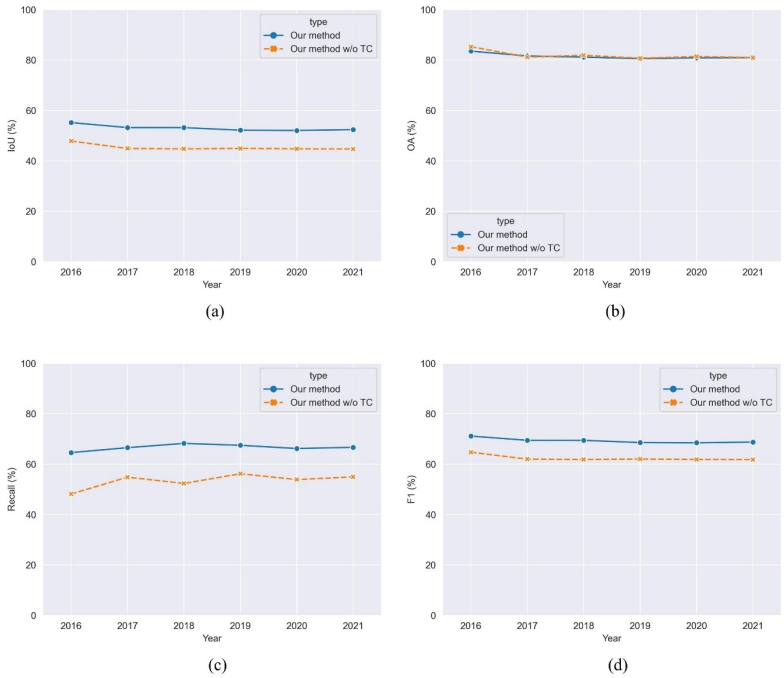


Figure S12: Temporal accuracy in Macao of our method and our method without TC loss. (a) IoU. (b) OA. (c) Recall. (d) F

Table S1: Statistics of building rooftop areas in the regions where the Sentinel-2 images are missing due to our cloud filter criteria. The image (cloud less than 20%) in 2021 is full coverage of China. Therefore, we utilize the CBRA results of 2021 as the benchmark to count the building rooftop area in the missing image region in each year (2016-2020) (the second column). In addition, we also calculate the proportion of missing building rooftop areas relative to the average area of building rooftops in China from 2016 to 2021 (89,826 Km^2) (the third column). It could be observed that the impact of the missing image is negligible.

	Building rooftop area in the region where the Sentinel-2 image is missing	The proportion of the missing building rooftop area
2016	704.89 Km^2	0.7847 %
2017	70.93 Km^2	0.0790 %
2018	36.82 Km^2	0.0410 %
2019	7.55 Km^2	0.0068 %
2020	5.90 Km^2	0.0066 %

Table S2: Descriptions of the built class in Dynamic World product, from its data report (Brown et al., 2022).

	Description	Example
Built area	• Clusters of human-made structures or individual very large human-made structures. • Contained industrial, commercial, and private buildings, and the associated parking lots. • A mixture of residential buildings, streets, lawns, trees, isolated residential structures or buildings surrounded by vegetative land covers. • Major road and rail networks outside of the predominant residential areas. • Large homogeneous impervious surfaces, including parking structures, large office buildings, and residential housing developments containing clusters of cul-de-sacs.	• Cluster of houses, can include smalls lawns or small patches of trees can be included • Dense villages, town, and cityscape (buildings and roads together) • Clusters of paved roads and large highways • Asphalt and other human-made surfaces

Table S3: The parameter list and our choice of the overall workflow.

Parameter	Description	Choice
Framework parameter		
H, W, C	The height, width and channel of the input image patch.	64, 64, 4 (i.e., RGB and NIR bands)
r	The up-sampling factor.	4
d	The dimension of the temporal representation.	128
τ	The temperature hyperparameter scaling the distribution of the TC loss.	0.75
N	The size of the memory bank maintaining the unrelated representation.	16
$\varepsilon = 10^{-6}$	Constant term used to make the Focal Tversky Loss numerically stable.	10-6
$\gamma = 0.5$	Focal term for shifting the model to converge to the easy samples (<1) or the hard samples (>1).	0.5
$\beta = 0.6$	Penalty term for the model focusing on minimizing FP predictions (<0.5) or the FN predictions (>0.5).	0.6
α: φ: ω	The ratio of the TC loss, HR loss and SG loss.	1:200:5
Training parameter		
Max epoch	The number of passes a training dataset	400
Learning rate	The updating pace of the parameter	0.00015
Optimizer	The mechanism to adjust the parameter	Adam optimizer with default setting
Augmentations	Rules for creating various images to tackle overfitting.	Flips, rotate and colour jittering
Batch size	The number of samples processed before the parameters updated	32 for high-resolution training set, 48 for low-resolution training set

Table S4: The impact of collaborative training with high-resolution references and low-resolution references. When using the SG loss, there are two strategies. One involves the use of low-resolution references solely in the absence of high-resolution supervised information, while the other entails training both simultaneously (i.e., utilizing low-resolution information for supervision at the same location even in the presence of high-resolution references). Our empirical results indicate that the second approach, i.e., collaborative training, can achieve superior accuracy compared to the former.

	IoU	OA	Recall	F1
STSR-Seg w/o collaborative training	45.15	82.73	74.51	62.21
STSR-Seg w/ collaborative training	45.51	82.85	74.66	62.55

References

Brown, C. F., Brumby, S. P., Guzder-Williams, B., Birch, T., Hyde, S. B., Mazzariello, J., Czerwinski, W., Pasquarella, V. J., Haertel, R., and Ilyushchenko, S.: Dynamic World, Near real-time global 10 m land use land cover mapping, Sci Data, 9, 1–17, 2022.

Malkin, K., Robinson, C., Hou, L., Soobitsky, R., Czawlytko, J., Samaras, D., Saltz, J., Joppa, L., and Jojic, N.: Label super-resolution networks, in: International Conference on Learning Representations, 2018.