論文 / 著書情報
Article / Book Information

論題(和文)	
Title(English)	ANNUAL PAST-PRESENT LAND COVER CLASSIFICATION FROM LANDSAT USING DEEP LEARNING FOR URBAN AGGLOMERATIONS
著者(和文)	CHINCHUTHAKUN, Varquez Alvin Christopher Galang, 山下 幸彦, 神田 学
Authors(English)	Worameth CHINCHUTHAKUN, David WINDERL, Alvin C.G. VARQUEZ, Yukihiko YAMASHITA, Manabu KANDA
出典(和文)	, Volume 12, Issue 2,
Citation(English)	Journal of JSCE, Volume 12, Issue 2,
発行日 / Pub. date	2024, 2
権利情報 / Copyright	本著作物の著作権は土木学会に帰属します。 Copyright (c) 2024 Japan Society of Civil Engineers.

ANNUAL PAST-PRESENT LAND COVER CLASSIFICATION FROM LANDSAT USING DEEP LEARNING FOR URBAN AGGLOMERATIONS

Worameth CHINCHUTHAKUN,[1] David WINDERL,[2] Alvin C.G. VARQUEZ,[3] Yukihiko YAMASHITA,[4] and Manabu KANDA[5]

[1]Dept. of Transdisciplinary Sci. and Eng., Tokyo Institute of Technology
(2-12-1 Ookayama, Meguro-ku, Tokyo, 152-8550, Japan)
E-mail: chinchuthakun.w.aa@m.titech.ac.jp (Corresponding Author)
[2]CIT Dept. of Computer Science, Technical University of Munich
(Boltzmannstraße 3 Garching, Germany)
E-mail: david.winderl@tum.de
[3]Member of JSCE, Associate Professor, Dept. of Transdisciplinary Sci. and Eng., Tokyo Institute of Technology
(2-12-1 Ookayama, Meguro-ku, Tokyo, 152-8550, Japan)
E-mail: varquez.a.aa@m.titech.ac.jp
[4]Professor, Center for Innovative Teaching and Learning, Tokyo Institute of Technology
(2-12-1 Ookayama, Meguro-ku, Tokyo, 152-8550, Japan)
[5]Member of JSCE, Professor, Dept. of Transdisciplinary Sci. and Eng., Tokyo Institute of Technology
(2-12-1 Ookayama, Meguro-ku, Tokyo, 152-8550, Japan)

Historical land cover data is crucial for understanding urbanization dynamics, climate modeling, and monitoring water resources. Following recent advancements in deep learning for processing Landsat archive data, prior studies have released high-resolution historical land cover maps on a global scale. However, these works often present prediction results limited to specific periods of coverage, which hinders their utility in conducting time series analysis across different urban agglomerations. To address this issue, we propose deep-learning models for land cover classification from Landsat images at a 30-meter spatial resolution. Our models are specifically designed for urban areas and are trained to be compatible with the sensors used in the Landsat series from 1972 to the present. Experimental results demonstrate that our models are highly effective in predicting land cover maps in new cities, particularly in built-up land and water regions. Our research provides pretrained models for land cover classification, facilitating future studies in related fields.

Key Words: *Landsat, land cover classification, deep learning, urban agglomerations*

1. INTRODUCTION

Land cover maps are essential for Earth surface monitoring and understanding the spatio-temporal distribution of different land cover types. Early global-scale land cover datasets include MODIS Land cover product[1)] and data from European Space Agency (ESA) Climate Change Initiative (abbreviated as ESA1992)[2)] at 500-m and 300-m spatial resolution, respectively. Following the advancement in remote sensing technology, higher resolution datasets have become available such as 10-m ESA WorldCover v200 product (abbreviated as ESA2021)[3)] for 2021 and 30-m Global Land Cover and Land Use (GLCLUC)[4)] for 2000 - 2020. However, these datasets may not be ideal for long-term analysis of urban land cover changes due to their limited temporal coverage. Hence, previous studies address this by training classifiers specifically for their study areas using these datasets instead.

Prior works generally employ traditional machine learning methods such as support vector machine (SVM)[5)], random forest[6)], and maximum likelihood estimation (MLE)[7)] due to their good trade-off between accuracy and computational costs. While these methods achieve high performance on a regional scale, they suffer from limited generalization capability when processing data from multiple urban agglomerations due to their pixel-level operation. This issue can be mitigated using deep-learning, which has been applied in some related works; however, the targets of these

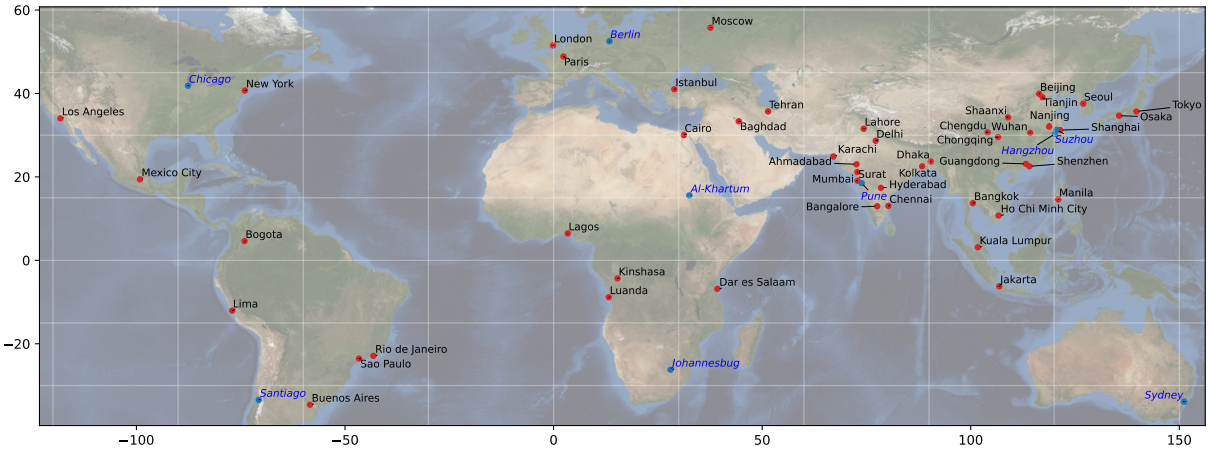


Fig. 1 Distribution of urban agglomerations for training (red) and testing (blue). Test cities consist of (1) the 5 largest non-megacity areas (Al-Khartum, Chicago, Hangzhou, Pune, and Suzhou) and (2) other 4 cities (Berlin, Johannesburg, Santiago, and Sydney).

works are limited either spatially[8] or temporally[9]. This study aims to (1) develop deep-learning models for 30-m land cover classification in urban agglomerations on a global scale and (2) demonstrate their potential as baselines for long-term analysis of urban land cover changes, particularly temporary or latest changes not yet available in existing datasets.

2. METHODOLOGY

In this section, we first discuss target areas including training data acquisition and its pre-processing. Then, we briefly describe network architecture, training loss, and implementation details.

(1) Training Data

We create a training set representing urban agglomerations using the 48 areas projected to become megacities in 2035, as declared by the United Nations[10]. As shown in Fig. 1, these cities cover different regions, ensuring diversity in urban landscape design. To mitigate geographical bias in the training data, the study area for each city is uniformly defined as a 40×40 km^2 around its center. For training, we use (1) ESA2021[3], selected for being one of the most accurate high-resolution (10-m) global-scale land cover datasets available, (2) atmospherically corrected surface reflectance data in the year 2021 from Landsat 7 and 8, and (3) digital number (DN) data during 1984 - 2013, which are available on Google Earth Engine[11].

After retrieving Landsat data, we apply bitmasks created from bits 3-5, 9, 11, and 13 in the Quality Assessment (QA_PIXEL) band to remove effects of medium and high confidence cloud, cloud shadow, and snow from the images. Then, we compute yearly median statistics and resample 30-m Landsat images to match with 10-m ESA2021 using nearest-neighbor interpolation. We prioritize using Tier 1 (T1) data because of their higher quality. Otherwise, we rely on Tier 2 (T2) images, particularly for older Landsat sensors due to data scarcity. While T2 data meet the radiometric standard of T1, they have less accurate geometry specifications due to factors such as imprecise orbital information and significant cloud cover.

(2) Network Architecture

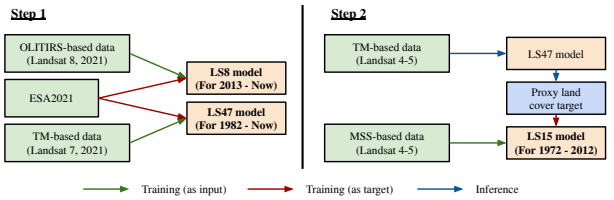


Fig. 2 Proposed training process. Predictions from LS47 are used to train LS15 because Landsat 5 ceased operation in 2013.

The Landsat series utilizes three different sensors: Multispectral Scanner System (MSS), Thematic Mapper (TM), and Operational Land Imager and Thermal Infrared Sensor (OLI-TIRS). Specifically, MSS is utilized in Landsat 1-5 (1972-2012), TM in Landsat 4-7 (1982-Now), and OLI-TIRS in Landsat 8 and beyond (2013-Now). We train separate models for each sensor to address the difference in band designations: LS15 for MSS, LS47 for TM, and LS8 for OLI-TIRS. As illustrated in Fig. 2, LS47 and LS8 are first trained in a supervised manner, using ESA2021 as reference data. Then, predictions from the LS47 are used as proxy targets when training LS15 by leveraging the fact that Landsat 4 and 5 are equipped with both MSS and TM sensors. Due to a lack of T1 images, we manually select high-quality T2 images for training LS15.

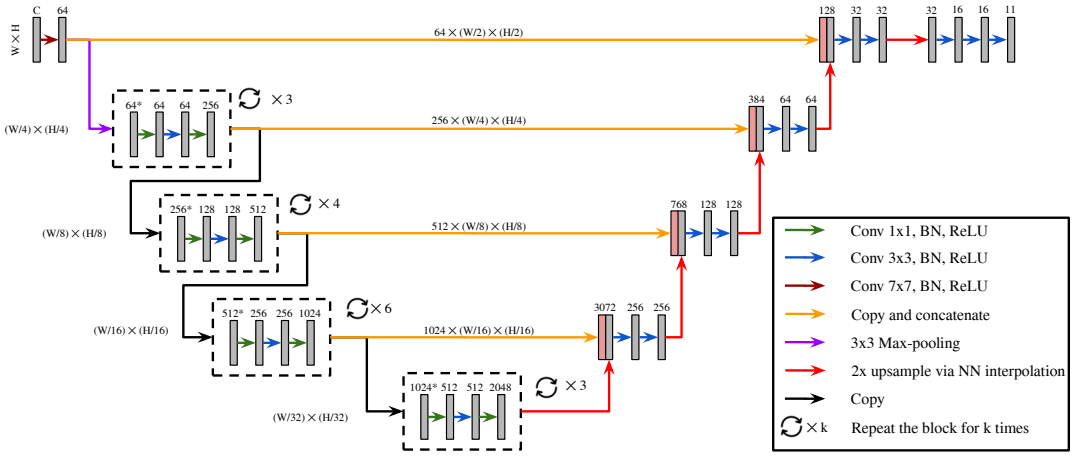


Fig. 3 Proposed network architecture based on U-Net with ResNet-50 as encoder. The grey and red blocks represent the feature map. The number above each block represents the number of feature channels it has. Input resolution is $(W, H) = (224, 224)$. The number of input channels C are 5, 10, and 11 for LS15, LS47, and LS8, respectively. BN denotes batch normalization[12].

As shown in Fig. 3, we adopt U-Net[13], one of standard network architectures with ResNet-50[14] as the encoder. To address the class imbalance issue, training loss is defined as:

$$L = L_1(\gamma_1) + \lambda L_2(\alpha, \beta, \gamma_2). \quad (1)$$

where L_1 and L_2 are Focal loss[15] and Focal Tversky Loss[16], respectively. After perform hyperparameter tuning via grid search, we use (1) focusing hyperparameters $\gamma_1 = \gamma_2 = 2.0$, (2) $\alpha = 0.3$ and $\beta = 0.7$ to balance the contributions of false positives and false negatives in L_2, and (3) $\lambda = 1.0$.

(3) Implementation Details

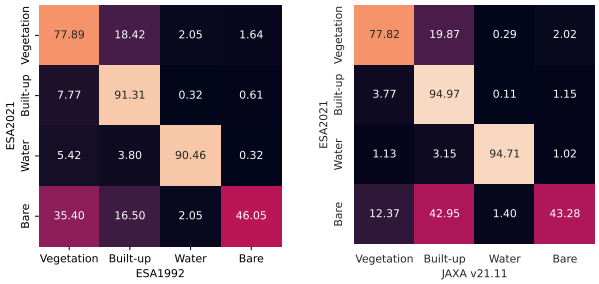
We implement the proposed models using PyTorch. For training, we use AdamW optimizer[17] with weight decay 1×10^{-5} and learning rate 5×10^{-4}. The learning rate is reduced by a factor of $\gamma = 0.5$ every 30 epochs. The total number of epochs is set to 100 with a batch size of 128. Each epoch consists of 50,000 image tiles of 224×224 pixels (equivalent to 2.24×2.24 km^2), which are normalized using means and standard deviation of train data. Besides employing standard data augmentation techniques, i.e. horizontal/vertical flipping and random patch swapping, we concatenate input tiles with additional three features: Normalized Difference Vegetation Index (NDVI), Normalized Difference Built-up Index (NDBI), and Built-up Index (BUI). Note that NDBI and BUI are excluded for LS15 because MSS lacks a short-wave infrared band.

3. EXPERIMENTS AND RESULTS

We present a series of experiments to assess the performance of the proposed models and investigate their potential applications. All experiments are conducted on a NVIDIA GeForce RTX 3090.

(1) Validation Data

Since the existing benchmark datasets designed for evaluating global land cover such as CGLS-LC100[18] contain few validation points in urban areas, we construct new validation datasets, shown in Table 1, based on available historical data to specifically validate the proposed models on urbanized regions. Excluding ESA2021-VAL, all datasets follow different class systems from that of ESA2021, requiring mapping functions to align them to common class labels before evaluation. Due to significant differences in the definition of several classes such as Cropland, Shrubland, and Grassland, we evaluate the model using 4 newly-defined super-classes: (1) Vegetation, (2) Built-up, (3) Water bodies, and (4) Bare or sparse vegetation.



(a) ESA2021 and ESA1992 in 2020, averaged over 48 train cities **(b)** ESA2021 and JAXA v21.11, averaged over Tokyo and Osaka

Fig. 4 Area-weighted confusion matrix of different datasets after label mapping and spatial resampling, normalized row-wise

Assuming that drastic land cover change within a few years is negligible, we inspect the annotation agreement between ESA2021 and ESA1992, as well as JAXA v21.11, after mapping the labels to the pro-

Table 1 Details of validation datasets derived from existing historical land cover data used in this paper. For each city indicated in Source, the study area is defined as 40×40 km^2 around its center. †: Land cover map represents the average situation in the period.

Dataset	Period of coverage	Spatial resolution	#Classes	#Validtion points	Source
ESA2021-VAL	2021	10 m	11	174,303,653	ESA2021,[3] at 9 test cities
ESA2021-PX	1984 - 1991	10 m	11	91,958,584	LS47 model, at Al-Khartum (1986), Chicago (1991), Hangzhou (1986), Pune (1989), and Suzhou (1984)
ESA1992-VAL	1992 - 2020	300 m	22	27,338,088	ESA1992,[2] at 48 train cities (annually)
JAXA16-VAL†	2006 - 2011	10 m	10	45,362,457	JAXA v16.09,[19] at Tokyo and Osaka
JAXA18-VAL†	2014 - 2016	30 m	10	5,044,620	JAXA v18.03,[19] at Tokyo and Osaka
JAXA21-VAL†	2018 - 2020	10 m	12	45,362,457	JAXA v21.11,[19] at Tokyo and Osaka

Table 2 Area-weighted producer's accuracy (PA) and user's accuracy (UA) of proposed models. LS8 and LS47 are evaluated on ESA2021-VAL while LS15 is validated on ESA2021-PX. Validation points do not contain Snow and Moss and lichen classes.

Model	Metric	Tree cover	Shrubland	Grassland	Cropland	Built-up	Bare	Water	Wetland	Mangroves
LS8	PA	0.786	0.137	0.567	0.716	0.855	0.641	0.979	0.146	0.764
	UA	0.810	0.459	0.580	0.660	0.813	0.579	0.976	0.168	0.527
LS47	PA	0.771	0.597	0.531	0.637	0.878	0.546	0.975	0.327	0.700
	UA	0.778	0.517	0.614	0.671	0.812	0.644	0.980	0.142	0.641
LS15	PA	0.489	0.066	0.545	0.704	0.599	0.894	0.959	0.072	–
	UA	0.605	0.082	0.589	0.527	0.823	0.726	0.941	0.231	–

posed system. As illustrated in Fig. 4, a significant inconsistency still remains for the Bare class in both datasets while Built-up and Water have an almost-perfect agreement. Thus, the analysis in these classes is expected to be highly reliable. We surmise that the class definitions are consistent across different JAXA datasets, although a direct comparison is not possible due to differences in the coverage period.

(2) Evaluation Metrics

Following previous studies[5),6),7)], we employ pixel-wise accuracy and the Kappa coefficient for evaluating models. For N-class classification problems, we denote $X = [x_{ij}]$ as a $N \times N$ confusion matrix such that x_{ij} represents the number of pixels classified as i-th and j-th class by the model and reference, respectively. Let $x_{+k} = \sum_i x_{ik}$, $x_{k+} = \sum_j x_{kj}$, and $s = \sum_{ij} x_{ij}$ for all $k \in C = \{1, \ldots, N\}$. Evaluation metrics can be computed as follows:

- **Overall accuracy**: $\mathrm{OA} = \mathbf{Tr}(X)/s$
- **Producer's accuracy**: $\mathrm{PA} = [x_{kk}/x_{+k}], \forall k \in C$
- **User's accuracy**: $\mathrm{UA} = [x_{kk}/x_{k+}], \forall k \in C$
- **Kappa coefficient**: $\mathrm{KC} = (p_0 - p_e)/(1 - p_e)$

where $p_0 = \mathbf{Tr}(X)/s$ and $p_e = \sum_k x_{+k}x_{k+}/s^2$.

(3) Generalization in New Urban Agglomerations

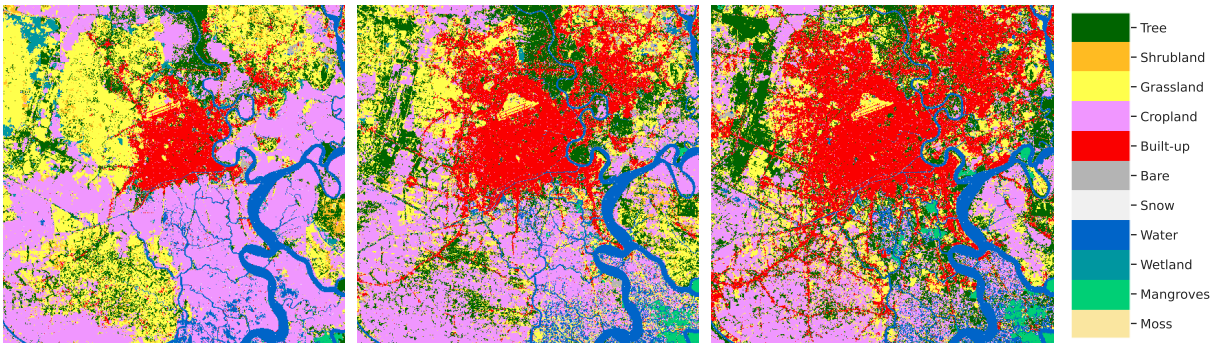
Using ESA2021-VAL as ground-truth, the proposed LS8 model achieves an OA of 0.771 and a KC of 0.711. LS47 shows competitive performance with identical OA and KC, but different PA and UA. As shown in Table 2, both models perform particularly well in the Built-up and Water bodies classes. Significant performance degradation is noticeable in minor classes such as Shurbland and Wetland.

(4) Effectiveness of Proxy Targets in LS15

Due to a lack of annual historical land cover data before 1992, we use land cover maps predicted by LS47 as reference data under the assumption that the difference between the real observations is negligible. Specifically, we evaluate LS15 using ESA2021-PX as ground-truth. Results in Table 2 suggest that LS15 trained with proxy targets performs fairly well in unseen cities, although its accuracy is limited by that of LS47. We found that land cover maps predicted by LS15 retain the overall outline but lack fine-grained details compared to LS47, which can be partially explained by the lower-quality and less informative bands in MSS compared to TM.

(5) Comparison with Historical Land Cover Data

We investigate the agreement of model predictions and historical land cover data on (1) a global scale (ESA1992-VAL) and (2) a more granular level, focusing on Tokyo and Osaka (JAXA16-VAL, JAVA18-VAL, and JAXA21-VAL). For validating with the former, we generate land cover maps for each city during 1992 - 2020 using LS47 and LS8 and downsample them spatially by computing mode statistics. For comparison with the latter, representative land cover maps are created using the pixel-wise mode from the down-



(a) 1995 (from LS47 with Landsat 5) **(b)** 2005 (from LS47 with Landsat 7) **(c)** 2015 (from LS8 with Landsat 8)

Fig. 5 Land cover change analysis at Ho Chi Minh City using LS47 and LS8. Model performance highly depends on the quality of input images. Noises in the prediction in year 2005 are partially caused by the incorrect Landsat data in some areas.

sampled model predictions. LS8 is used during 2014 - 2020 because of higher Landsat data availability. Fig. 5 shows land cover changes in Ho Chi Minh City.

Table 3 indicates that predictions from our models highly correlate with the historical data from ESA1992-VAL, achieving an OA of 0.833 and a KC of 0.738. However, accuracy gradually degrades as the year deviates from 2021 (i.e. the year of training data) as shown in Fig. 6. Similar trends in JAXA datasets are observed in Table 3. One possible explanation is covariate shift such as temporal differences in urban landscape design and lower-quality Landsat data.

(6) Application in Hydrology

Unlike existing datasets with a fixed temporal resolution, our models can be utilized to investigate temporary land cover changes caused by specific incidents, such as floods. We conducted a case study focusing on the heavily submerged areas during the 2022 Pakistan floods near Larkana. We predict a land cover map using the 40×40 km^2 Landsat 8 data from June to October 2022 (preprocessed as described in Section 2.1). As shown in Fig. 7, the water regions identified using LS8 closely resemble areas with relatively high Normalized Difference Water Index (NDWI), suggesting that the models are applicable to this task despite not being explicitly trained to detect floods. The inconsistency between the predicted mask and NDWI can be partially attributed to significant cloud cover.

4. CONCLUSION

This work presents an approach or tool that can automatically generate historical annual land cover maps for any specific urbanized region of interest. This approach is composed of interdependent deep-learning models, specifically trained for urban agglomerations and all sensors employed in the Landsat series. Our

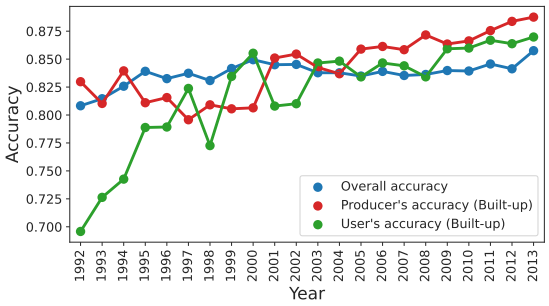


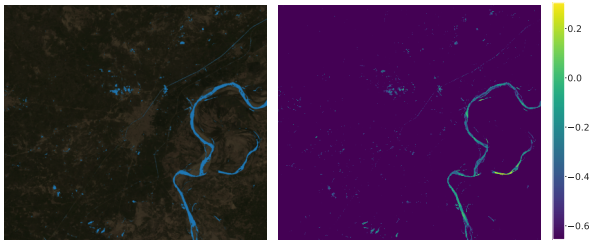
Fig. 6 Annual area-weighted overall accuracy and Built-up accuracy of LS47 evaluated on ESA1992 from 1992 - 2014

Table 3 Area-weighted producer's accuracy (PA) and user's accuracy (UA) of LS8 and LS47 evaluated on historical land cover datasets. ESA1992 results are temporally averaged. Bare class is excluded due to low inter-dataset agreement.

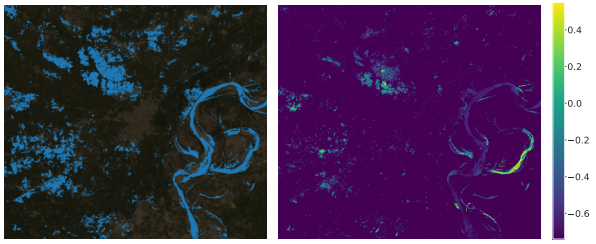
Dataset	Metric	Vegetation	Built-up	Water
ESA1992-VAL	PA	0.838	0.858	0.899
	UA	0.828	0.828	0.907
JAXA21-VAL	PA	0.790	0.966	0.978
	UA	0.884	0.914	0.958
JAXA18-VAL	PA	0.857	0.957	0.930
	UA	0.828	0.920	0.984
JAXA16-VAL	PA	0.677	0.985	0.923
	UA	0.926	0.863	0.973

experiments demonstrated their applicability in new cities across different areas and the high pixel-wise correspondence between generated land cover maps and existing historical data.

The main limitation of this study is that the performance of models heavily relies on the input image quality, and the presented experiments only measure the accuracy of the generated land cover maps concerning ESA2021 and other historical datasets. Since these datasets are also subjected to classification errors during construction, the results do not necessarily reflect the accuracy with respect to the actual ground-truth. Moreover, as the models are trained to produce ESA2021-like predictions, their theoretical accuracy



(a) Left: ESA2021, Right: NDWI in 2021 (Prior to flooding)



(b) Left: Prediction from LS8, Right: NDWI (June - October 2022)

Fig. 7 Comparison of water masks obtained from land cover maps and NDWI values (clipped to [-0.1, 1.0] for visualization)

with respect to the real ground-truth cannot exceed that of ESA2021. Exploring new network architectures or training methods for learning from noisy labels could be promising directions for future research.

In future work, we plan to evaluate the models on more urban agglomerations and historical datasets with less annotation disagreement. We also aim to calibrate the models to provide interpretable probabilities as output and further simplify API, enhancing their usefulness in applications for non-machine learning proficient users.

ACKNOWLEDGMENTS: This work was supported by JSPS KAKENHI Grant Numbers JP21H04573 and JP21K14249. Data has been provided through ALOS of the Japan Aerospace Exploration Agency, ESA CCI Land Cover project, and the European Space Agency.

REFERENCES

1) Sulla-Menashe, D., Gray, J. M., Abercrombie, S. P., & Friedl, M. A. (2019). Hierarchical mapping of annual global land cover 2001 to present: The modis collection 6 land cover product. *Remote Sensing of Environment*, *222*, 183–194. https://doi.org/10.1016/j.rse.2018.12.013
2) Copernicus Climate Change Service. (2019). Land cover classification gridded maps from 1992 to present derived from satellite observations. https://doi.org/10.24381/CDS.006F2C9A
3) Zanaga, D., Van De Kerchove, R., Daems, D., De Keersmaecker, W., *et al.* (2022). *Esa worldcover 10 m 2021 v200*. Zenodo. https://doi.org/10.5281/zenodo.7254221
4) Potapov, P., Hansen, M. C., Pickens, A., Hernandez-Serna, A., *et al.* (2022). The global 2000-2020 land cover and land use change dataset derived from the landsat archive: First results. *Frontiers in Remote Sensing*, *3*. https://doi.org/10.3389/frsen.2022.856903
5) Ul Din, S., & Mak, H. W. L. (2021). Retrieval of land-use/land cover change (lucc) maps and urban expansion dynamics of hyderabad, pakistan via landsat datasets and support vector machine framework. *Remote Sensing*, *13*(16). https://doi.org/10.3390/rs13163337
6) Souverijns, N., Buchhorn, M., Horion, S., Fensholt, R., *et al.* (2020). Thirty years of land cover and fraction cover changes over the sudano-sahel using landsat time series. *Remote Sensing*, *12*(22). https://doi.org/10.3390/rs12223817
7) Mancino, G., Falciano, A., Console, R., & Trivigno, M. L. (2023). Comparison between parametric and non-parametric supervised land cover classifications of sentinel-2 msi and landsat-8 oli data. *Geographies*, *3*(1), 82–109. https://doi.org/10.3390/geographies3010005
8) Tzepkenlis, A., Marthoglou, K., & Grammalidis, N. (2023). Efficient deep semantic segmentation for land cover classification using sentinel imagery. *Remote Sensing*, *15*(8). https://doi.org/10.3390/rs15082027
9) Karra, K., Kontgis, C., Statman-Weil, Z., Mazzariello, J. C., *et al.* (2021). Global land use / land cover with sentinel 2 and deep learning. *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*, 4704–4707. https://doi.org/10.1109/IGARSS47720.2021.9553499
10) United Nations, Departments of Economic and Social Affairs, Population Division. (2019). World Urbanization Prospects: The 2018 Revision. *Online Edition*.
11) Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., *et al.* (2017). Google earth engine: Planetary-scale geospatial analysis for everyone [Big Remotely Sensed Data: tools, applications and experiences]. *Remote Sensing of Environment*, *202*, 18–27. https://doi.org/10.1016/j.rse.2017.06.031
12) Ioffe, S., & Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. *CoRR*, *abs/1502.03167*. http://arxiv.org/abs/1502.03167
13) Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, 234–241.
14) He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770–778. https://doi.org/10.1109/CVPR.2016.90
15) Lin, T.-Y., Goyal, P., Girshick, R., He, K., *et al.* (2017). Focal loss for dense object detection. *2017 IEEE International Conference on Computer Vision (ICCV)*, 2999–3007. https://doi.org/10.1109/ICCV.2017.324
16) Abraham, N., & Khan, N. M. (2019). A novel focal tversky loss function with improved attention u-net for lesion segmentation. *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, 683–687. https://doi.org/10.1109/ISBI.2019.8759329
17) Loshchilov, I., & Hutter, F. (2017). Fixing weight decay regularization in adam. *CoRR*, *abs/1711.05101*. http://arxiv.org/abs/1711.05101
18) Tsendbazar, N., Herold, M., Li, L., Tarko, A., *et al.* (2021). Towards operational validation of annual global land cover maps. *Remote Sensing of Environment*, *266*, 112686. https://doi.org/https://doi.org/10.1016/j.rse.2021.112686
19) Japan Aerospace Exploration Agency, Earth Observation Research Center. (2022). High-Resolution Land Use and Land Cover Map of Japan. https://www.eorc.jaxa.jp/ALOS/en/dataset/lulc_e.htm

(Received May 31, 2023)
(Accepted September 12, 2023)