https://doi.org/10.5194/essd-2022-75
Preprint. Discussion started: 22 June 2022

Open Access Earth System Science Data Discussions

UGS-1m: Fine-grained urban green space mapping of 34 major cities in China based on the deep learning framework

Qian Shi[1,2], Mengxi Liu[1,2,*], Andrea Marinoni[3,4], and Xiaoping Liu[1,2]

[1]School of Geography and Planning, Sun Yat-sen University, Guangzhou 510275, China
[2]Guangdong Key Laboratory for Urbanization and Geo-simulation, Guangzhou 510275, China
[3]Dept. of Physics and Technology, UiT the Arctic University of Norway, 9019 Tromsø, Norway
[4]Dept. of Engineering, University of Cambridge, Cambridge CB2 1PZ, UK

Correspondence: Mengxi Liu (liumx23@mail2.sysu.edu.cn)

Abstract. Urban green space (UGS) is an important component in the urban ecosystem and has great significance to the urban ecological environment. Although the development of remote sensing platforms and deep learning technologies have provided opportunities for UGS mapping from high-resolution images (HRIs), challenges still exist in its large-scale and fine-grained application, due to insufficient annotated datasets and specially designed methods for UGS. Moreover, the domain shift between images from different regions is also a problem that must be solved. To address these issues, a general deep learning (DL) framework is proposed for UGS mapping in the large scale, and the fine-grained UGS maps of 34 major cities/areas in China are generated (UGS-1m). The DL framework consists of a generator and a discriminator. The generator is a fully convolutional network designed for UGS extraction (UGSNet), which integrates attention mechanisms to improve the discrimination to UGS, and employs a point rending strategy for edge recovery. The discriminator is a fully connected network aiming to deal with the domain shift between images. To support the model training, an urban green space dataset (UGSet) with a total number of 4,454 samples of size 512×512 is provided. The main steps to obtain UGS-1m can be summarized as follows: a) Firstly, the UGSNet will be pre-trained on the UGSet in order to get a good starting training point for the generator; b) After pre-training on the UGSet, the discriminator is responsible to adapt the pre-trained UGSNet to different cities/areas through adversarial training; c) Finally, the UGS results of the 34 major cities/areas in China (UGS-1m) are obtained using 2,343 Google Earth images with a data frame of 7’30" in longitude and 5’00" in latitude, and a spatial resolution of nearly 1.1 meters. Evaluating the performance of the proposed approach on samples from Guangzhou city shows the validity of the UGS-1m products, with an overall accuracy of 87.4% and an F1 score of 81.14%. Furthermore, experiments on UGSet with the existing state-of-the-art (SOTA) DL models proves the effectiveness of UGSNet, with the highest F1 of 77.30%. Finally, the comparisons with existing products further shows the feasibility of the UGS-1m and the effectiveness and great potential of the proposed DL framework. The UGS-1m can be downloaded from https://doi.org/10.5281/zenodo.6155516 (Shi et al., 2022).

1 Introduction

Urban green space (UGS), one of the most important components of the urban ecosystem, refers to the vegetation entity in the urban area (Kuang and Dou, 2020), such as parks and green buffers. It plays a very important role in the urban ecological

https://doi.org/10.5194/essd-2022-75
Preprint. Discussion started: 22 June 2022


Open Access Earth System Science Data Discussions

environment (Kong et al., 2014; Zhang et al., 2015), public health (Fuller et al., 2007) and social economy (De Ridder et al., 2004). However, in the context of rapid urbanization, UGS is now facing drastic changes in terms of sustainability, integrity and diversity. Indeed, some UGS changes might have a negative impact on the fragile urban ecosystem and cause a series of problems on the urban environment and welfare. In technical literature, several works have enthusiastically discussed the interaction mechanism between UGS and urban environment, in which relevant UGS data is indispensable (see for instance Zhou and Wang, 2011; Huang et al., 2018; Zhao et al., 2010). Thus, to provide the reliable basic geographic data for in-depth UGS research, fast and accurate mapping of UGS is crucial and necessary.

With the development and application of remote sensing technology, diversified remote sensing data are increasingly used to obtain UGS coverage. In this respect, multispectral remote sensing images are widely used. Sun et al. (2011) extracted UGS in China's 117 metropolises from MODIS data over the last three decades through Normalized Difference Vegetation Index (NDVI), to study its impacts on urbanization. Huang et al. (2017) obtained urban green coverage of 28 megacities from Landsat images between 2005 and 2015 to assess the change of health benefits by urban green spaces. Recently, taking advantage of cloud computing, many excellent land cover products based on Landsat and Sentinel-1&2 images have been proposed, including GlobeLand30 (Jun et al., 2014), GLC_FCS30 (Zhang et al., 2021) , FROM_GLC10 (Gong et al., 2013), Esri 2020 LC (Helber et al., 2019). These products have provided valuable world-wide maps of land coverage, so that researchers can easily extract relevant information and conduct in-depth research on specific UGS properties, such as impervious surface, UGS coverage, etc. Although multispectral images have provided powerful data support for large-scale and long-term UGS monitoring, it is often difficult to obtain UGS information of small scale due to the limitation of spatial resolution of multispectral images. In other words, some small-scale UGSs (such as UGS attached to buildings and roads) are difficult to be identified in multispectral images, although they are of great significance to urban ecosystem. Therefore, images with higher spatial resolution are required to address the large difference in intra-class scale of urban green space.

To get finer-grained extraction of UGS, remote sensing imagery with richer spatial information are more and more employed in UGS extraction, such as Rapid-Eye, ALOS and SPOT images (Mathieu et al., 2007; Zhang et al., 2015; Zhou et al., 2018). In these studies, machine learning methods, including SVM (Yang et al., 2014) and random forest (Huang et al., 2017), are often employed to obtain UGS coverage. However, hand-craft features are required for classification in these methods, which are time- and labor-consuming, and not objective enough.

Deep learning (DL) based methods can hence be used to address these issues (Deng and Yu, 2014). In fact, DL schemes can extract multi-level features automatically, so that they are becoming the mainstream solution in many fields, including computer vision, natural language processing, medical image recognition, etc (Zhang et al., 2018; Devlin et al., 2018; Litjens et al., 2017). Among DL algorithms, the full convolution networks (FCNs), represented by UNet (Ronneberger et al., 2015), SegNet (Badrinarayanan et al., 2017) and Deeplab v3+ (Chen et al., 2018a), have been widely introduced into remote sensing interpretation tasks, such as building footprint extraction (Liu et al., 2019a), change detection (Liu et al., 2021), as well as UGS mapping (Liu et al., 2019b). For instance, Xu et al. (2020) improved the U-Net model by adding batch normalization (BN) and dropout layer to solve the over-fitting problem, and monitored UGS areas in Beijing. Liu et al. (2019b) employed DeepLab v3+ to automatically obtain green space distribution from GF-2 imagery. With the help of convolutional operators

https://doi.org/10.5194/essd-2022-75
Preprint. Discussion started: 22 June 2022


Open Access Earth System Science Data Discussions

with different receptive fields for multi-scale feature extraction and fully convolutional layers to recover spatial information, the FCN methods can achieve accurate pixel-level results in an end-to-end manner (Daudt et al., 2018).

In the context of rapid changes in the global ecological environment, large-scale and high-resolution automatic extraction of UGS is becoming more and more important (Cao and Huang, 2021; Wu et al., 2021). Although the existing methods have achieved good results in UGS extraction based on deep learning, there are still open problems to be solved. Firstly, significant intra-class differences and inter-class similarities of UGS have jeopardized the classic strategies for recognition of UGS. The appearance and scale of UGS vary significantly due to the wide variety involved. For example, while the green buffers inside roads are measured in meters, a public park could be measured in kilometers. Moreover, the substantial similarity between farmland and UGS also leads to severe misclassification, while farmland does not belong to UGS. Therefore, guaranteeing that the model can extract effective relevant features is crucial to accurately obtain UGS coverage.

Secondly, the development of UGS extraction methods based on deep learning framework is greatly limited by the lack of datasets, while accurate and reliable results by deep learning models heavily rely on sufficient training samples. The last few decades have witnessed the flourishing of many large datasets to be used for deep learning architectures, such as ImageNet (Krizhevsky et al., 2012), PASCAL VOC (Everingham et al., 2015), SYSU-CD (Shi et al., 2021). Nevertheless, due to tremendous time and labor required, there are few publicly available datasets with fine-grained UGS information. This condition reduces the efficiency of researchers, and hinders the fair comparison between UGS extraction methods, not to mention providing reliable basic data for large-scale UGS mapping.

Last but not least, the large-scale fine-grained UGS mapping is also limited by the difference of data distribution. Affected by external factors (e.g., illumination, angle and distortion), remote sensing images collected in different regions and time are difficult to keep consistent data distribution. Therefore, the model trained on a certain dataset fail to be well applied to images of another region. In order to overcome the data shift between difference data, domain adaptation should be adopted to improve the generalization of the model.

In view of the aforementioned problems, we develop a deep learning framework for large-scale and high-precision UGS extraction, leading to a collection of 1-meter UGS products of 34 major cities/areas in China (UGS-1m). As shown in Figure 1, we firstly construct a high-resolution urban green space dataset (UGSet), which contains 4,454 samples of size 512×512, to support training and verification of UGS extraction model. Then we build a deep learning model for UGS mapping, which consists of a generator and a discriminator. The generator is a fully convolutional network for UGS extraction, also referred as UGSNet, which integrates an enhanced Coordinate attention (ECA) module to capture more effective feature representations, and a point head module to get fine-grained UGS results. The discriminator is a fully connected network that aims to adapt the UGSet-pretrained UGSNet to large-scale UGS mapping through adversarial training (Tsai et al., 2018). Finally, the UGS results of the 34 major cities/areas in China, namely UGS-1m, are obtained after post-processing, including mosaic and mask.

The contributions of this paper can be summarized as follows:

(1) a general deep learning framework for large-scale and high-resolution UGS mapping, and generating UGS maps of 34 major cities/areas in China with a spatial resolution of 1 meter (UGS-1m), is proposed;

https://doi.org/10.5194/essd-2022-75
Preprint. Discussion started: 22 June 2022

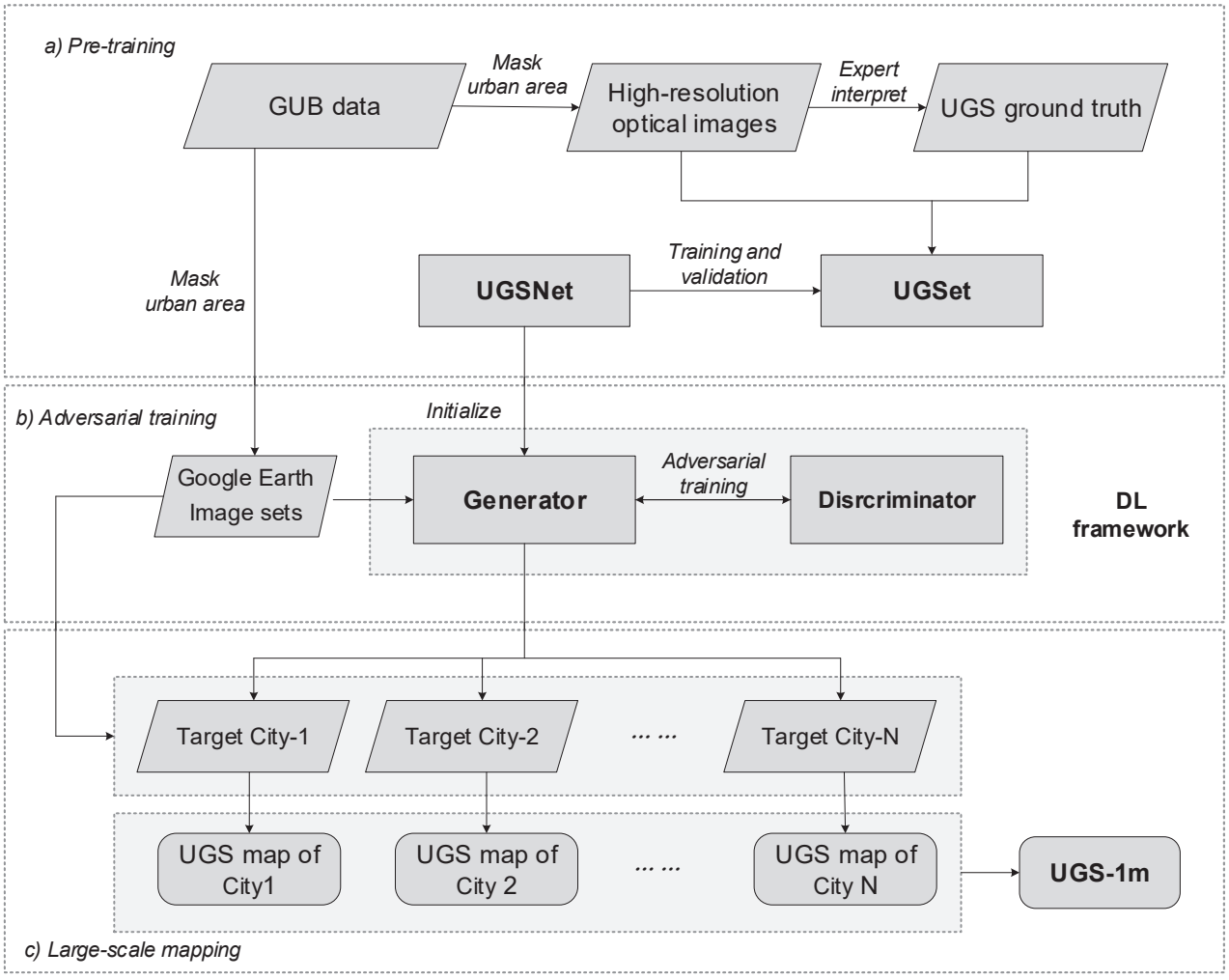


Figure 1. Diagram of the deep learning framework to generate UGS-1m.

(2) a fully convolutional network for fine-grained UGS mapping (UGSNet) is introduced. This architecture integrates an enhanced Coordinate attention (ECA) module and a point head module to address the intra-class differences and inter-class similarities in UGS;

(3) a large benchmark dataset, Urban Green Space dataset (UGSet), is provided to support and foster the UGS research based on the deep learning framework;

The reminder of this paper is arranged as follows. Sect. 2 introduces the study area and data. Sect. 3 illustrates the deep learning framework for UGS mapping. Sect. 4 assesses and demonstrates the UGS results. Then discussions will be conducted in Sect. 5. Finally, conclusions will be made in Sect. 6.

https://doi.org/10.5194/essd-2022-75
Preprint. Discussion started: 22 June 2022

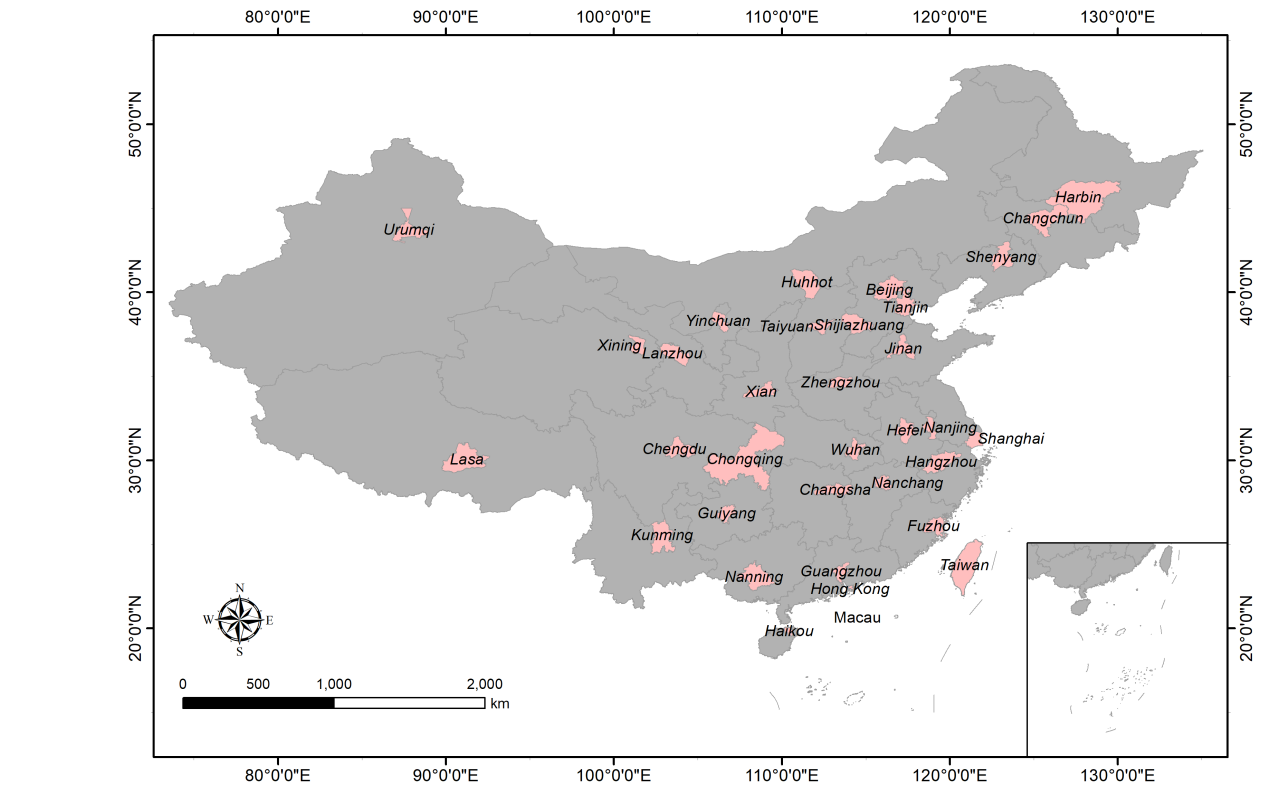


Figure 2. The 34 major cities/areas in China.

2 Study area and data

2.1 Study area

As Figure 2 shows, the study area is 34 major cities/areas in China, including two special administrative regions (Hong Kong and Macau), four municipalities (Beijing, Shanghai, Tianjin, and Chongqing), capitals of five autonomous regions (Huhuhot, Nanning, Lasa, Yinchuan and Urumqi), as well as captials of 23 provinces (Harbin, Changchun, Shenyang, Shijiazhuang, Lanzhou, Xining, Xi'an, Zhengzhou, Jinan, Changsha, Wuhan, Nanjing, Chengdu, Guiyang, Kunming, Hangzhou, Nanchang, Guangzhou, Fuzhou, Taipei, Haikou).

According to the newly promulgated "Guidance of the General Office of the State Council on Scientific Greening", it is required to scientifically promote urban gardening and greening construction; improve the urban green space system and reasonable layout of greening land to meet the requirements of urban health, safety and livability; promote urban ecological restoration, central city, old city greenway construction, to facilitate green travel for urban residents; build a balanced distribution of the park system. Thus, strengthening the management of urban landscaping and the design and construction

https://doi.org/10.5194/essd-2022-75
Preprint. Discussion started: 22 June 2022


Open Access Earth System Science Data Discussions

Table 1. UGS types and descriptions.

Type	Description
Park	Green space open to the public for all kinds of outdoor activities
Green buffer	Green space to isolate facilities like sewage treatment plants, garbage treatment plants, high-voltage lines, etc, or water bodies such as rivers, lakes and seas
Square green space	Green space in open scape area with leisure and entertainment functions, mainly is shrubs and grassland
Attached green space	Green space attached to residential, transportation, industrial, or commercial land
Other green space	Green space on undefined land

management of major greening projects, scientific proof and implementation, is necessary to ensure the quality of the project, and to build a livable environment in which people and nature live in harmony.

2.2 Datasets

2.2.1 UGSet

Urban green space can be divided into five categories, including park, green buffer, square green space, attached green space and other green space (Chen et al., 2018b), as described in Table 1. Different types of UGS vary not only on their functions, but also on shape and scale: these properties become more apparent in high-resolution images. For instance, park and green buffer are often occurring in a relatively large volume, while attached green space and square green space are mainly scattered in urban areas in smaller form. In other words, urban green space is not only diverse, but also has large inter- and intra-class scale differences. Therefore, a dataset that contains UGS samples of different types and scales is an important guarantee for the model to learn and identify UGS accurately.

In order to provide an extensive sample database for wide-range UGS mapping, as well as a benchmark for comparisons among deep learning algorithms, we constructed a largescale high-resolution urban green space dataset (UGSet), which contains 4,544 images of size 512×512 with a spatial resolution of nearly 1 meter. These images acquired by Gaofen-2 are collected from 142 images in Guangdong Province, China, as shown in Figure 3. With the aim to filter out green space in non-urban areas, the global urban boundaries (GUB) data (Li et al., 2020) of 2018 is used to mask the urban areas of each original image. All types of UGS in the images are carefully annotated through expert visual interpretation, before they are cropped into 512×512 patches. As can be seen from Figure 4, the category of non-UGS and UGS in the ground truth are represented by 0 and 255, respectively. According to the ratio of 5:2:3, the UGSet is randomly divided into the training set, verification set and test set.

https://doi.org/10.5194/essd-2022-75
Preprint. Discussion started: 22 June 2022

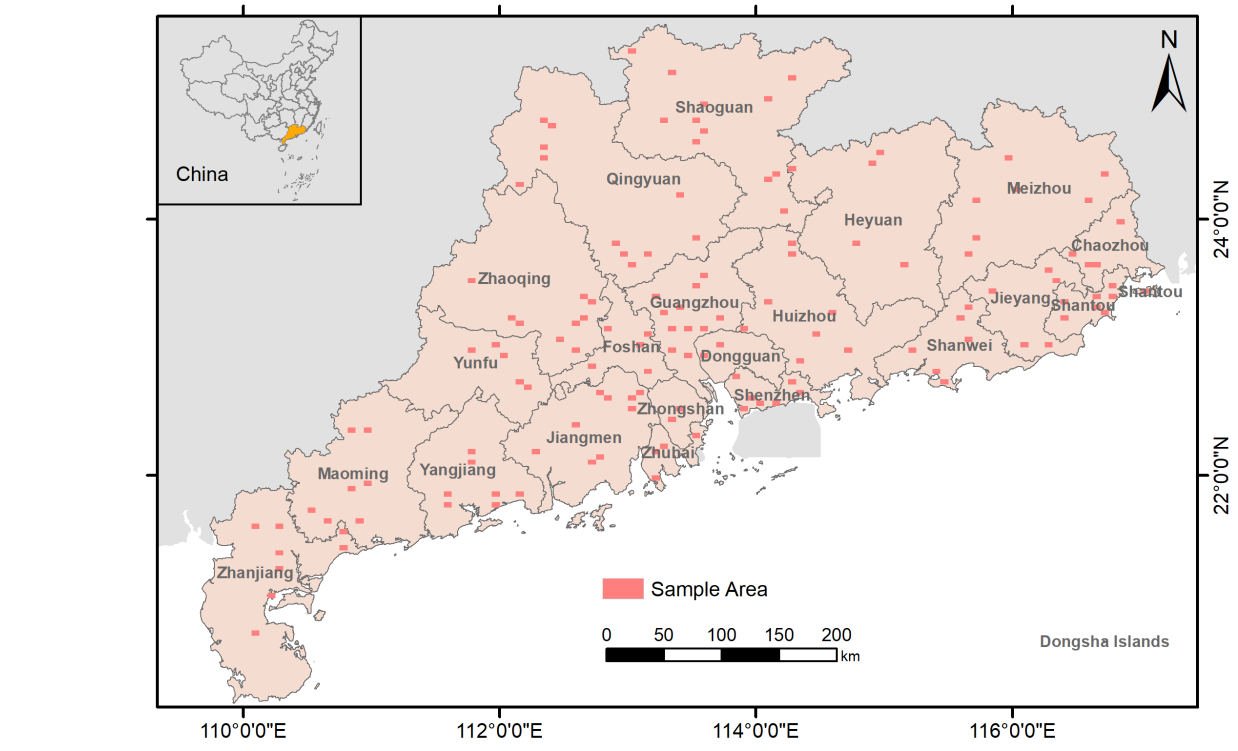


Figure 3. The 142 Sample areas in UGSet collected from Guangdong Province. Each with a data frame of 3’45 "in longitude and 2’30" in latitude.

2.2.2 Global urban boundaries (GUB)

The global urban boundaries (GUB) data (Li et al., 2020) that delineate the boundary of global urban area in seven years (i.e. 1990, 1995, 2000, 2005, 2010, 2015, and 2018) is obtained by processing the 30 m global artificial impervious area (GAIA) data (Gong et al., 2020). It is worth noting that GAIA is the only annual map of impervious surface areas from 1985 to 2018 with a resolution of 30 m. In this study, the GUB data in 2018 are adopted to mask the urban area. Specifically, in order to obtain accurate UGS samples, the GUB data are used to filter out non-relevant green space samples from non-urban areas. The GUB data are also applied to the UGS results from the model for post processing, so to get final UGS map of each city/area.

2.2.3 Google Earth Imagery

Google Earth is a free software which enables users to view high-resolution satellite images around the world. Therefore, in order to obtain fine-grained UGS maps in the study area, a total number of 2,343 Google Earth images covering 34 major cities/areas in China are downloaded, each with a data frame of 7’30" in longitude and 5’00" in latitude, and a spatial resolution

https://doi.org/10.5194/essd-2022-75
Preprint. Discussion started: 22 June 2022

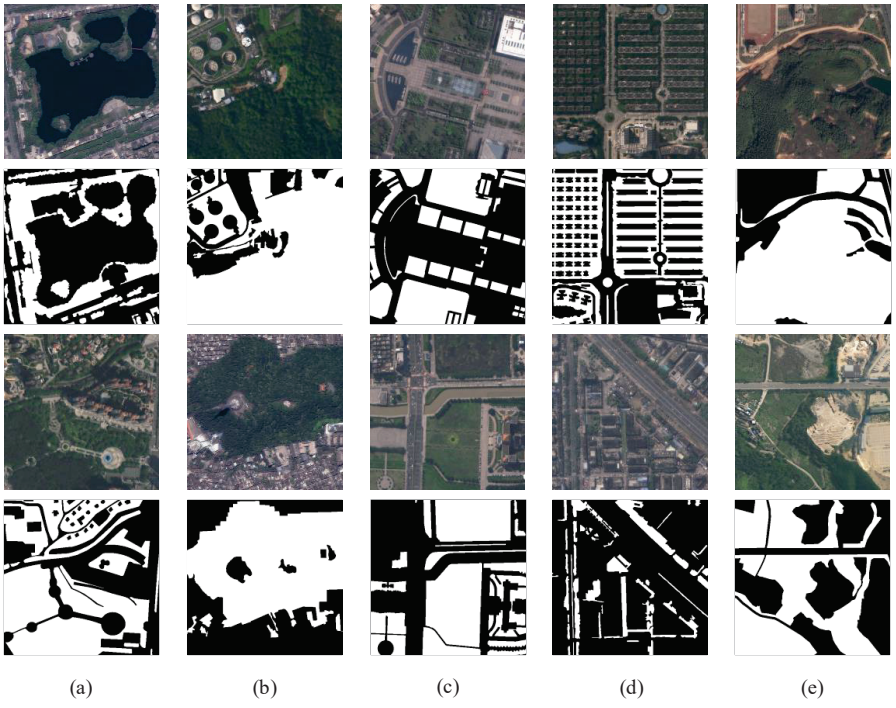


Figure 4. Example of different types of UGS samples in UGSet (Images were retrieved from GF-2 2019). (a) Park; (b) Green buffer; (c) Square green space; (d) Attached green space; (e) Other green space.

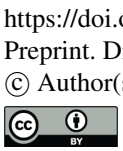
of nearly 1.1 meters. All images selected are clear and cloud-free to avoid missed detection. Limited by the GPU memory, these images are all cropped into the size of 512×512 for prediction.

3 Methods

In order to realize large-scale fine-grained UGS mapping, a general model framework is essential, in addition to a sufficiently large dataset. Therefore, we propose a deep learning framework for UGS mapping: its functional flowchart is shown in Figure 5. Inspired by adversarial domain adaptation frameworks (Tsai et al., 2018), the proposed framework includes a generator and a discriminator. In particular, a fully convolutional neural network, namely UGSNet, is designed as the generator: this structure is utilized to learn and extract fine-grained UGS information. On the other hand, a simple fully connected network is employed as the discriminator to help model domain adaptation and achieve large-scale UGS mapping.

The following Sect. 3.1 and Sect. 3.2 will introduce the structure of UGSNet and discriminator, respectively. The optimization process of the deep learning framework will be described in Sect. 3.3, which can be divided into two parts: pre-training and adversarial training. Parameter settings and accuracy evaluation will be covered in Sect. 3.4 and Sect. 3.5.

https://doi.org/10.5194/essd-2022-75
Preprint. Discussion started: 22 June 2022



Open Access Earth System Science Data Discussions

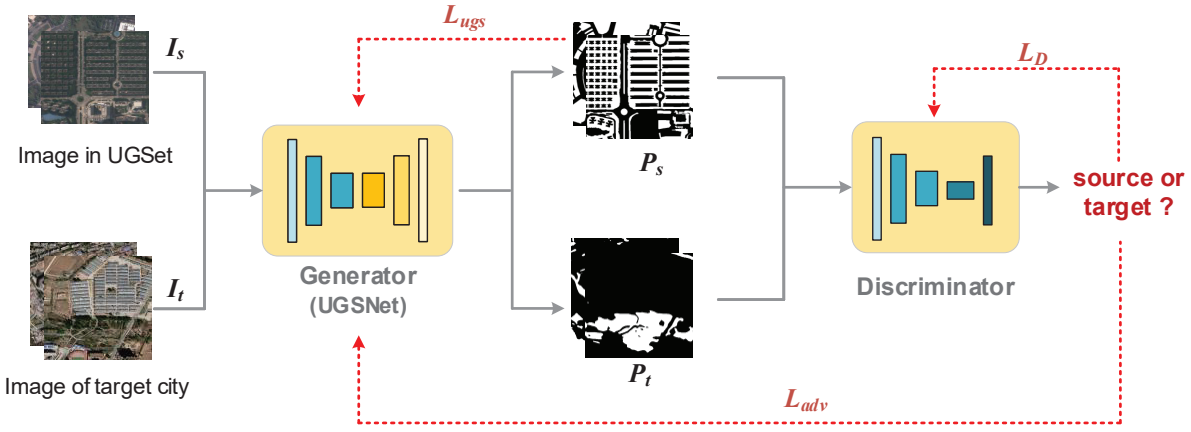


Figure 5. Flowchart of the proposed deep learning framework for UGS mapping (The "Image in UGSet" were retrieved from GF-2 2019, whiel the "Image of target city" © Google Earth 2020).

3.1 UGSNet

As shown in Figure 6, UGSNet contains two parts: a backbone to extract multi-scale features and generate coarse results, and a point head module to obtain fine-grained results.

3.1.1 Backbone

The backbone of the UGSNet first adopts the efficient ResNet-50 as feature extractor to capture multi-scale features from the images. This segment contains five stages: the first stage consists of a 7×7 convolutional layer, a batch normalization layer (Ioffe and Szegedy, 2015), a Rectified Linear Unit (ReLU) function (Glorot et al., 2011) and a max-pooling layer with a stride of 2; then, four residual blocks are utilized to capture deep features of four different levels. The four residual blocks are connected by four enhanced Coordinate attention (ECA) modules (Hou et al., 2021) to enhance feature representations. The structure of the ECA module is shown in Figure 7.

Then, four 1×1 convolutional blocks will be applied to the attention-refined features of the four residual blocks to unify their output channels to 96, then they are concatenated together after resizing. Finally, the fused features will be input into two 3×3 convolutional layers to generate a coarse prediction map, which is 1/4 the size of the input image.

3.1.2 Point Head

The point head introduces the point rending strategy (Kirillov et al., 2020) to get fine-grained UGS results. Specifically, given the coarse UGS results from the backbone, the point head will firstly collect N sampling points. Then, the N points are mapped to the coarse UGS results as well as the selected fine-grained features from the backbone (the output of Res-Block2 in this

https://doi.org/10.5194/essd-2022-75
Preprint. Discussion started: 22 June 2022

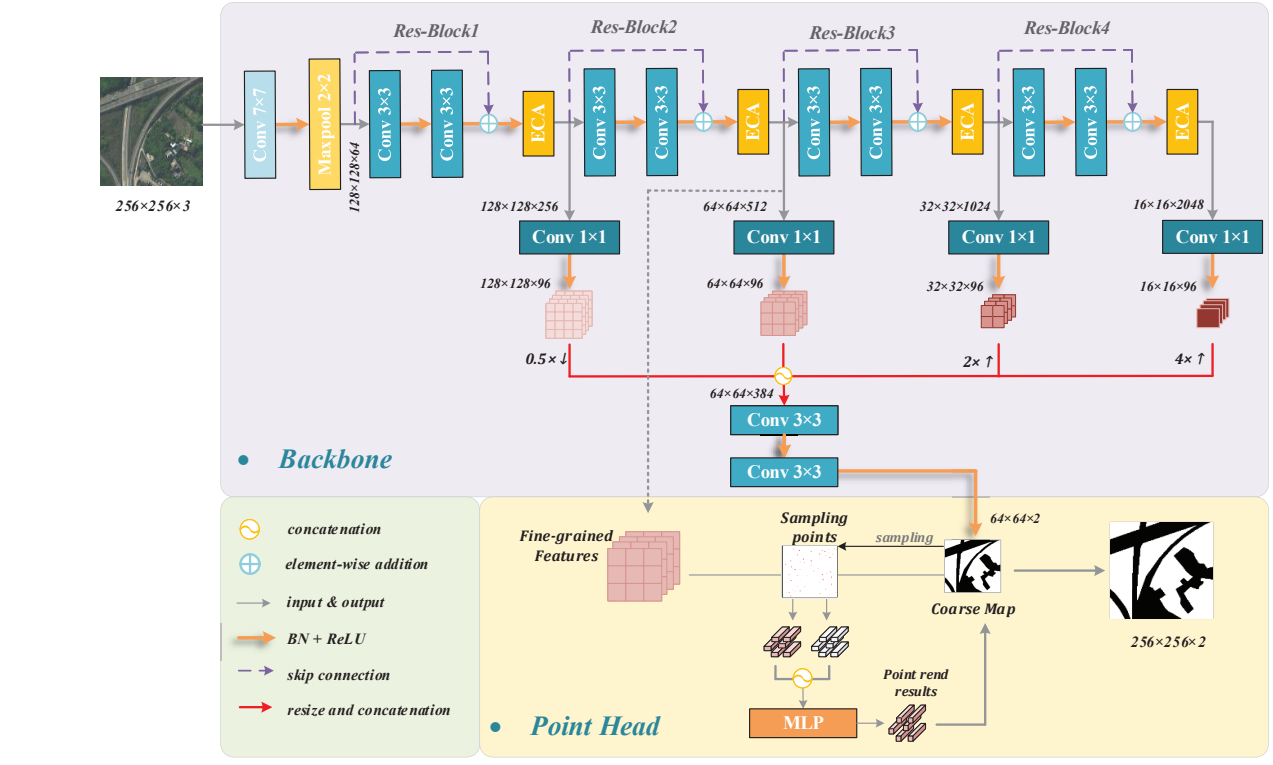


Figure 6. Architecture of the proposed UGSNet (The image was retrieved from GF-2 2019).

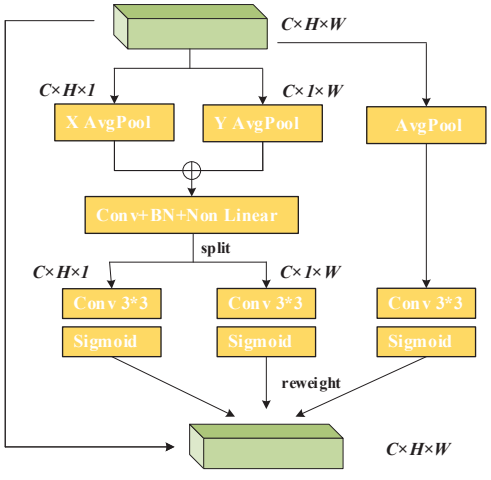


Figure 7. Structure of the enhanced Coordinate attention (ECA) module.

https://doi.org/10.5194/essd-2022-75
Preprint. Discussion started: 22 June 2022

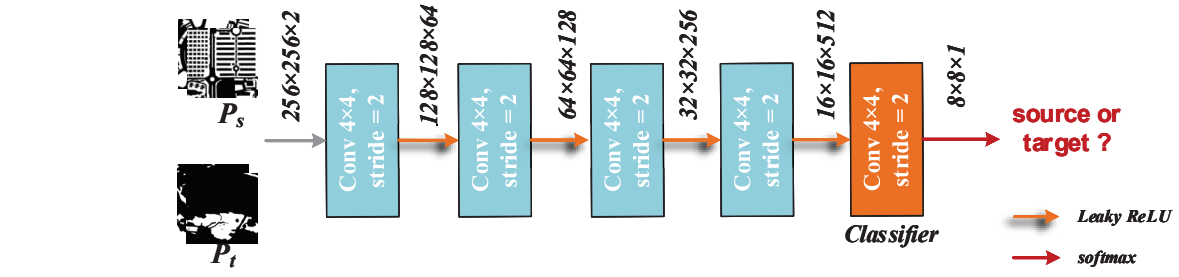


Figure 8. The structure of the discriminator.

paper) to extract the feature vector of each sampling point. Then, the point-wise features will be concatenated and input into a simple multilayer perceptron (MLP) to obtain classification results of each point.

3.2 Discriminator

In order to transfer the prior knowledge from UGSet to images from other regions, a discriminator is adopted to obtain a well-adapt UGSNet for each city/area in an unsupervised way. As shown in Figure 8, the discriminator consists of five convolutional layers with a kernel size of 4 and a stride of 2, each connected by a Leaky ReLU layer. The output channels of each convolutional layers are 64, 128, 256, 512, and 1, respectively. Finally, a softmax layer is utilized to conduct binary classification based on the output feature.

3.3 Optimization

The training of the proposed deep learning framework can be divided into two steps: pre-training, and adversarial training. At the beginning, the UGSNet will be fully trained on UGSet to get initial parameters for the generator. After that, the discriminator will be adopted to help generalize the pre-trained UGSNet to target cities/areas through adversarial learning. Detail information of the optimization process are described in the following.

3.3.1 Pre-training

In the pre-training process, the UGSNet will learn characteristics of all kinds of UGS from UGSet. Let us suppose the coarse result output by the backbone is X and the ground truth is Y. Then, the loss between Y and X is calculated by a dice loss, which can be defined as follows:

$$L_{Dice} = 1 - (2|X \cap Y|)/(|X| + |Y|) \quad (1)$$

where $|X \cap Y|$ is the intersection between X and Y, whilst $|X|$ and $|Y|$ denote the number of elements of X and Y, respectively.

https://doi.org/10.5194/essd-2022-75
Preprint. Discussion started: 22 June 2022


Open Access Earth System Science Data Discussions

The loss of the classification results of the N sampling points in the point head is measured by the cross-entropy loss, which can be defined as

$$L_{CE} = -\sum_{i}^{N}[x_i logy_i + (1-x_i)log(1-y_i)] \tag{2}$$

wherex_i and y_i represent the results and ground truth of the i-th point among the N sampling ones, respectively.

Finally, the UGSNet is optimized by a hybrid loss, which can be expressed by

$$L_{ugs} = L_{Dice} + L_{CE} \tag{3}$$

3.3.2 Adversarial training

After pre-training, the UGSNet is employed as the generator in the deep learning framework and train with the discriminator to obtain a model that can be used for the UGS extraction of a target city/area. Taking the image I_s and ground truth Y_s in UGSet, and the image I_t from the target city/area as input, the adversarial training process requires no additional data for supervision, which can be summarized as follows:

(1) Taking the pre-trained UGSNet as the start training point of the generator, the I_s and I_t are forward to the generator to get their prediction result P_s and P_t;

(2) Input P_s and P_t into the discriminator in turn to distinguish the source of the inputs;

(3) According to the judgement result, the discriminator will be optimized first, which can be denoted as

$$L_D(P_s, P_t) = -[(1-y)log(D(P_s)^{(h,w,0)}) + ylog(D(P_t)^{(h,w,1)})] \tag{4}$$

where y represents the source of the inputs, and $y = 0$ denotes an input from I_t, and $y = 1$ denotes an input from I_s.

(4) Then, an adversarial loss L_adv is calculated to help promote the generator to produce more similar results to confuse the discriminator. The L_adv is actually the loss when the discriminator misclassifies the source of P_t as I_s, which can be expressed as

$$L_{adv} = -log(D(P_t)^{(h,w,1)}) \tag{5}$$

(5) Finally, the generator will be optimized through the following objective function:

$$L_G = L_{ugs} + L_{adv} \tag{6}$$

3.4 Parameter settings

During the pre-training process, the training set of the UGSet is used for parameter optimization, to which random crop, flip and rotation are employed for data augmentation to avoid overfitting, while the verification set was used to monitor the training

https://doi.org/10.5194/essd-2022-75
Preprint. Discussion started: 22 June 2022


Open Access Earth System Science Data Discussions

direction and save the model in time. Five common semantic segmentation models are selected for comparison to prove the validity of UGSNet, including UNet (Ronneberger et al. 2015), SegNet (Badrinarayanan et al. 2017), UperNet (Xiao et al. 2018), BiSeNet (Yu et al. 2018) and PSPNet (Zhao et al. 2017). In addition, ablation study is also conducted to further verify the effectiveness of the ECA modules and the point head. All models are fully trained for 200 epochs based on Adam optimizer with an initial learning rate of 0.0001, which begins to decline linearly in the last 100 epochs. A batch size of 8 sample pairs is adopted. After training, all selected models were compared through their performance on the test set. The adversarial training process lasts for 10000 epochs, in which the batch size is set to 2. Both the generator and the discriminator employ an initial learning rate of 0.0001. All experiments are implemented in PyTorch environments and are conducted on the GeForce RTX 2080ti to accelerate model training

3.5 Accuracy Evaluation

Five indices are involved in the evaluation, including precision (Pre), recall (Rec), F1-score, intersection-over-union (IoU), and overall accuracy (OA). Given that TP, FP, TN and FN refer to true positives, false positives, true negatives, and false negatives, respectively, these indices can be defined as follows

$$Pre = \frac{TP}{TP + FP} \tag{7}$$

$$Rec = \frac{TP}{TP + FN} \tag{8}$$

$$F1 = \frac{2precision \cdot recall}{precision + recall} \tag{9}$$

$$IoU = \frac{TP}{FP + TP + FN} \tag{10}$$

$$OA = \frac{TP + TN}{FP + TP + FN + TN} \tag{11}$$

During the pre-training process, the Pre, Rec, F1 and IoU are utilized to measure the model performance on UGSet, which are commonly used in semantic segmentation tasks. On the other and, the Pre, Rec, F1 and OA indices are employed to verify the accuracy of the generated UGS maps (UGS-1m).

4 Results

To evaluate the reliability of the UGS results, four example areas are randomly collected in Guangzhou city, as shown in Figure 10-(e), so as to obtain their UGS ground truth as the reference maps by expert interpretation. Then, accuracy evaluation

https://doi.org/10.5194/essd-2022-75
Preprint. Discussion started: 22 June 2022

Open Access Earth System Science Data Discussions

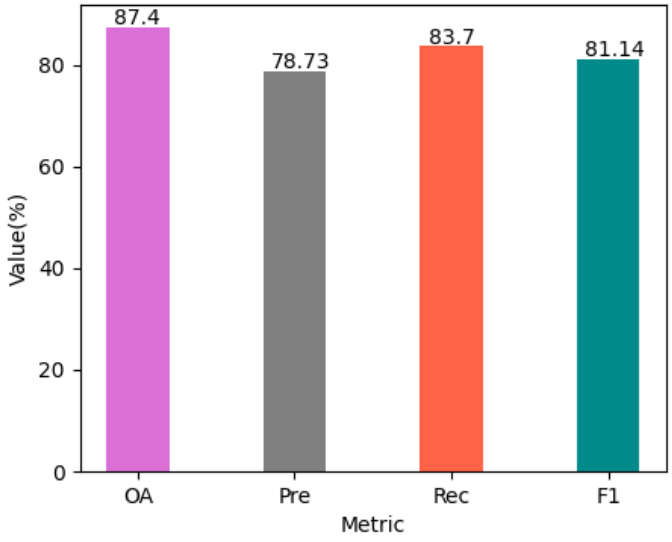


Figure 9. Results of accuracy evaluation on UGS-1m.

is conducted according to the reference maps and results in UGS-1m. These results are summarized in Figure 9. It is worth noting that the OA of the UGS results can reach 87.4%, and the F1 can reach 81.14%. Moreover, the 83.7% Recall also indicates a relatively low missed-detection rate of the UGS extraction results, which is significantly important in operational scenarios. The quantitative results preliminarily demonstrate the validity of UGS-1m products.

Visual comparisons between the reference map and the UGS-1m results are displayed in (a)-(d) of Figure 10. From the overview image of (a)-(d), it can be seen that the results in UGS-1m are in good agreement with the reference map, which is mainly reflected in the good restoration of UGS of various scales in each example image. The zoom-in area of each image further shows the details of UGS-1m for extracting different kinds of UGS, including park, square, green buffer, as well as the attached green space. Notably, the UGS-1m performs well in the extraction of green space attached to residential buildings, although they are complex and broken in morphology compared to other UGS types. The visualization further illustrates the reliability of the UGS-1m. All results in the UGS-1m are demonstrated in Figure 11.

5 Discussions

5.1 Comparison with deep learning models

As Table 2 shows, the proposed method outperforms all state-of-the-art (SOTA) baselines with the highest F1 and IoU of 77.30% and 62.99%, respectively. The scheme that achieves the closest performance, PSPNet, obtains an IoU of 60.96%. The ablation study indicates that the integration of the ECA modules and point head can improve the Base model by 0.17% and

https://doi.org/10.5194/essd-2022-75
Preprint. Discussion started: 22 June 2022

Open Access Earth System Science Data Discussions

Table 2. Performance of different semantic segmentation methods on UGSet.

Method	Pre(%)	Rec(%)	F1(%)	IoU(%)
SegNet	71.70	77.34	74.42	59.26
UNet	75.57	74.92	75.25	60.31
UperNet	74.81	76.11	75.45	60.58
BiSeNet	75.13	76.26	75.69	60.89
PSPNet	76.57	74.94	75.74	60.96
Base	76.52	76.22	76.37	61.77
Base+attention	**76.87**	76.13	76.49	61.94
Base+PointHead	74.99	78.89	76.89	62.46
UGSNet	75.40	**79.29**	**77.30**	**62.99**

0.69% on IoU, respectively, which proves their effectiveness. The IoU of UGSNet is 1.22% higher than that of the Base model, indicating that their combination has a greater gain effect. Figure 12 further demonstrates the performance of different methods. Especially, the UGSNet also does a best job in the mapping of many small scale UGS, as shown in the yellow boxes in Figure 12.

5.2 Comparison with existing products

We compare the UGS-1m results with existing global land use products to verify the reliability of the results, including GlobeLand30 (Chen and Chen, 2018), GLC_FCS30 (Zhang et al., 2021) and Esri 2020 LC (© 2021 Esri). Due to different classification systems, these products need to be reclassified in two categories first. Specifically, forests, grasslands and shrublands are reclassified as UGS, while the other categories are reclassified as non-UGS. Examples from 6 different cities of different latitudes, including Changchun, Urumqi, and Beijing, Chengdu, Wuhan and Guangzhou, are collected to give more comprehensive demonstrations on our UGS results. The visualization comparison among UGS-1m, GlobeLand30, GLC_FCS30 and Esri 2020 LC is shown in Figure 13. Apparently, the three comparative products contain most large-scale UGS, among which the GlobeLand30 performs best with most complete UGS prediction. However, many detailed UGS features are still missed due to the limitation on spatial resolution of source image. On the other hand, the UGS-1m provides UGS with relatively large scale, as well as detailed UGS information such as attached green space. The results and comparisons fully demonstrate the effectiveness and potential of the proposed deep learning framework for large-scale and fine-grained UGS mapping.

5.3 Limitations and future work

At present, the availability of high-resolution images is still severely limited by factors such as temporal resolution, image distortion and cloud occlusion. Therefore, the Google Earth images used to produce UGS-1m are very difficult to collect at one time, so it is difficult to ensure the unity of phenology. In the proposed deep learning framework, we introduce domain adaptation to deal with this problem to some extent. Comparison results had shown the effectiveness of UGS-1m, as well as the

https://doi.org/10.5194/essd-2022-75
Preprint. Discussion started: 22 June 2022

feasibility and potential of the proposed deep learning framework for large-scale, high-resolution mapping. Future works will be dedicated to extract UGS information based on data with higher temporal resolution, such as SAR images and unmanned aerial vehicle (UAV) images.

6 Code and data availability

The UGS-1m product can be downloaded at https://doi.org/10.5281/zenodo.6155516 (Shi et al., 2022). They are named by name of the 34 cities/areas.

The Dataset and Code for the deep learning framework will be available at https://github.com/liumency/UGS-1m.

7 Conclusions

In this paper, we propose a novel deep learning (DL) framework for large-scale UGS mapping, and generate the fine-grained UGS maps for 34 major cities/area in China (UGS-1m). The accuracy evaluation on the UGS-1m products indicates the reliability and applicability. Comparative experiments conducted on UGSet among several SOTA semantic segmentation networks show that UGSNet can achieve the best performance on UGS extraction. The ablation study on UGSNet also demonstrates the effectiveness of the ECA module and point head. Comparisons between UGS-1m and existing land use products have proved the validity of the proposed DL framework for large-scale and fine-grained UGS mapping. The achievements provided in this paper can support the scientific community for UGS understanding and characterization, and pave the way for the development of robust and efficient methods able to tackle the current limits and needs of UGS analysis in technical literature.

Author contributions. Qian Shi: Conceptualization, Resources, Investigation, Funding acquisition and Writing – review & editing; Mengxi Liu: Data curation, Methodology, Validation, Formal analysis, Visualization and Writing – original draft preparation; Andrea Marinoni: Writing – review & editing; Xiaoping Liu: Writing – review & editing.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. This study is supported in part by the National Natural Science Foundation of China under Grant 61976234, in part by Centre for Integrated Remote Sensing and Forecasting for Arctic Operations (CIRFA) and the Research Council of Norway (RCN Grant no. 237906), and the Visual Intelligence Centre for Research-based Innovation funded by the Research Council of Norway (RCN Grant no. 309439).

https://doi.org/10.5194/essd-2022-75
Preprint. Discussion started: 22 June 2022

References

Badrinarayanan, V., Kendall, A., and Cipolla, R.: Segnet: A deep convolutional encoder-decoder architecture for image segmentation, IEEE transactions on pattern analysis and machine intelligence, 39, 2481–2495, https://doi.org/https://doi.org/10.1109/TPAMI.2016.2644615, 2017.

Cao, Y. and Huang, X.: A deep learning method for building height estimation using high-resolution multi-view imagery over urban areas: A case study of 42 Chinese cities, Remote Sensing of Environment, 264, 112 590, 2021.

Chen, J. and Chen, J.: GlobeLand30: Operational global land cover mapping and big-data analysis, Science China. Earth Sciences, 61, 1533–1534, 2018.

Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., and Adam, H.: Encoder-decoder with atrous separable convolution for semantic image segmentation, 2018a.

Chen, W., Huang, H., Dong, J., Zhang, Y., Tian, Y., and Yang, Z.: Social functional mapping of urban green space using remote sensing and social sensing data, ISPRS Journal of Photogrammetry and Remote Sensing, 146, 436–452, 2018b.

Daudt, R. C., Saux, B. L., and Boulch, A.: Fully Convolutional Siamese Networks for Change Detection, IEEE, 2018.

De Ridder, K., Adamec, V., Ba N Uelos, A., Bruse, M., B U Rger, M., Damsgaard, O., Dufek, J., Hirsch, J., Lefebre, F., P E Rez-Lacorzana, J. M., and Others: An integrated methodology to assess the benefits of urban green space, Science of the total environment, 334, 489–497, https://doi.org/https://doi.org/10.1016/j.scitotenv.2004.04.054, 2004.

Deng, L. and Yu, D.: Deep Learning: Methods and Applications, Foundations & Trends in Signal Processing, 7, 197–387, 2014.

Devlin, J., Chang, M., Lee, K., and Toutanova, K.: BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, CoRR, abs/1810.04805, http://arxiv.org/abs/1810.04805, 2018.

Everingham, M., Eslami, S. A., Van Gool, L., Williams, C. K., Winn, J., and Zisserman, A.: The pascal visual object classes challenge: A retrospective, International journal of computer vision, 111, 98–136, https://doi.org/https://doi.org/10.1007/s11263-014-0733-5, 2015.

Fuller, R. A., Irvine, K. N., Devine-Wright, P., Warren, P. H., and Gaston, K. J.: Psychological benefits of greenspace increase with biodiversity, Biology letters, 3, 390–394, https://doi.org/https://doi.org/10.1098/rsbl.2007.0149, 2007.

Glorot, X., Bordes, A., and Bengio, Y.: Deep Sparse Rectifier Neural Networks, Journal of Machine Learning Research, 15, 315–323, 2011.

Gong, P., Wang, J., Yu, L., Zhao, Y., Zhao, Y., Liang, L., Niu, Z., Huang, X., Fu, H., Liu, S., Li, C., Li, X., Fu, W., Liu, C., Xu, Y., Wang, X., Cheng, Q., Hu, L., Yao, W., Zhang, H., Zhu, P., Zhao, Z., Zhang, H., Zheng, Y., Ji, L., Zhang, Y., Chen, H., Yan, A., Guo, J., Yu, L., Wang, L., Liu, X., Shi, T., Zhu, M., Chen, Y., Yang, G., Tang, P., Xu, B., Giri, C., Clinton, N., Zhu, Z., Chen, J., and Chen, J.: Finer resolution observation and monitoring of global land cover: first mapping results with Landsat TM and ETM+ data, International journal of Remote Sensing, 34, 2607–2654, https://doi.org/10.1080/01431161.2012.748992, 2013.

Gong, P., Li, X., Wang, J., Bai, Y., Chen, B., Hu, T., Liu, X., Xu, B., Yang, J., Zhang, W., and Zhou, Y.: Annual maps of global artificial impervious area (GAIA) between 1985 and 2018, Remote Sensing of Environment, 236, 111 510, https://doi.org/https://doi.org/10.1016/j.rse.2019.111510, 2020.

Helber, P., Bischke, B., Dengel, A., and Borth, D.: Eurosat: A novel dataset and deep learning benchmark for land use and land cover classification, IEEE journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2019.

Hou, Q., Zhou, D., and Feng, J.: Coordinate Attention for Efficient Mobile Network Design, 2021.

Huang, C., Yang, J., Lu, H., Huang, H., and Yu, L.: Green spaces as an indicator of urban health: evaluating its changes in 28 mega-cities, Remote Sensing, 9, 1266, https://doi.org/https://doi.org/10.3390/rs9121266, 2017.

https://doi.org/10.5194/essd-2022-75
Preprint. Discussion started: 22 June 2022

Huang, C., Yang, J., and Jiang, P.: Assessing impacts of urban form on landscape structure of urban green spaces in China using Landsat images based on Google Earth Engine, Remote Sensing, 10, 1569, https://doi.org/https://doi.org/10.3390/rs10101569, 2018.

Ioffe, S. and Szegedy, C.: Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift, CoRR, abs/1502.03167, http://arxiv.org/abs/1502.03167, 2015.

Jun, C., Ban, Y., and Li, S.: China: Open access to Earth land-cover map, Nature, 514, 434–434, https://doi.org/https://doi.org/10.1038/514434c, 2014.

Kirillov, A., Wu, Y., He, K., and Girshick, R.: PointRend: Image Segmentation As Rendering, in: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 9796–9805, https://doi.org/10.1109/CVPR42600.2020.00982, 2020.

Kong, F., Yin, H., James, P., Hutyra, L. R., and He, H. S.: Effects of spatial pattern of greenspace on urban cooling in a large metropolitan area of eastern China, Landscape and Urban Planning, 128, 35–47, https://doi.org/https://doi.org/10.1016/j.landurbplan.2014.04.018, 2014.

Krizhevsky, A., Sutskever, I., and Hinton, G. E.: Imagenet classification with deep convolutional neural networks, Advances in neural information processing systems, 25, 1097–1105, 2012.

Kuang, W. and Dou, Y.: Investigating the patterns and dynamics of urban green space in China's 70 major cities using satellite remote sensing, Remote Sensing, 12, 1929, https://doi.org/https://doi.org/10.3390/rs12121929, 2020.

Li, X., Gong, P., Zhou, Y., Wang, J., Bai, Y., Chen, B., Hu, T., Xiao, Y., Xu, B., Yang, J., Liu, X., Cai, W., Huang, H., Wu, T., Wang, X., Lin, P., Li, X., Chen, J., He, C., Li, X., Yu, L., Clinton, N., and Zhu, Z.: Mapping global urban boundaries from the global artificial impervious area (GAIA) data, Environmental Research Letters, 15, 094 044, https://doi.org/10.1088/1748-9326/ab9be3, 2020.

Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., van der Laak, J. A., van Ginneken, B., and Sánchez, C. I.: A survey on deep learning in medical image analysis, Medical Image Analysis, 42, 60–88, https://doi.org/https://doi.org/10.1016/j.media.2017.07.005, 2017.

Liu, M., Shi, Q., Marinoni, A., He, D., and Zhang, L.: Super-resolution-based Change Detection Network with Stacked Attention Module for Images with Different Resolutions, 2021.

Liu, P., Liu, X., Liu, M., Shi, Q., Yang, J., Xu, X., and Zhang, Y.: Building Footprint Extraction from High-Resolution Images via Spatial Residual Inception Convolutional Neural Network, Remote Sensing, 11, 2019a.

Liu, W., Yue, A., Shi, W., Ji, J., and Deng, R.: An Automatic Extraction Architecture of Urban Green Space Based on DeepLabv3plus Semantic Segmentation Model, 2019b.

Mathieu, R., Aryal, J., and Chong, A. K.: Object-based classification of Ikonos imagery for mapping large-scale vegetation communities in urban areas, Sensors, 7, 2860–2880, https://doi.org/https://doi.org/10.3390/s7112860, 2007.

Ronneberger, O., Fischer, P., and Brox, T.: U-net: Convolutional networks for biomedical image segmentation, pp. 234–241, 2015.

Shi, Q., Liu, M., Li, S., Liu, X., Wang, F., and Zhang, L.: A deeply supervised attention metric-based network and an open aerial image dataset for remote sensing change detection, IEEE Transactions on Geoscience and Remote Sensing, https://doi.org/https://doi.org/10.1109/TGRS.2021.3085870, 2021.

Shi, Q., Liu, M., and Marinoni, A.: UGS-1m: Fine-grained urban green space mapping of 34 major cities in China based on the deep learning framework [Data set], Zenodo, https://doi.org/10.5281/zenodo.6155516, 2022.

Sun, J., Wang, X., Chen, A., Ma, Y., Cui, M., and Piao, S.: NDVI indicated characteristics of vegetation cover change in China's metropolises over the last three decades, Environmental monitoring and assessment, 179, 1–14, https://doi.org/https://doi.org/10.1007/s10661-010-1715-x, 2011.

https://doi.org/10.5194/essd-2022-75
Preprint. Discussion started: 22 June 2022

Tsai, Y.-H., Hung, W.-C., Schulter, S., Sohn, K., Yang, M.-H., and Chandraker, M.: Learning to Adapt Structured Output Space for Semantic Segmentation, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018.

Wu, F., Wang, C., Zhang, H., Li, J., Li, L., Chen, W., and Zhang, B.: Built-up area mapping in China from GF-3 SAR imagery based on the framework of deep learning, Remote Sensing of Environment, 262, 112 515, https://doi.org/https://doi.org/10.1016/j.rse.2021.112515, 2021.

Xu, Z., Zhou, Y., Wang, S., Wang, L., Li, F., Wang, S., and Wang, Z.: A novel intelligent classification method for urban green space based on high-resolution remote sensing images, Remote Sensing, 12, 3845, https://doi.org/https://doi.org/10.3390/rs12223845, 2020.

Yang, J., Huang, C., Zhang, Z., and Wang, L.: The temporal trend of urban green coverage in major Chinese cities between 1990 and 2010, Urban Forestry & Urban Greening, 13, 19–27, https://doi.org/https://doi.org/10.1016/j.ufug.2013.10.002, 2014.

Zhang, B., Li, N., Wang, S., and Others: Effect of urban green space changes on the role of rainwater runoff reduction in Beijing, China, Landscape and Urban Planning, 140, 8–16, https://doi.org/https://doi.org/10.1016/j.landurbplan.2015.03.014, 2015.

Zhang, Q., Yang, L. T., Chen, Z., and Li, P.: A survey on deep learning for big data, Information Fusion, 42, 146–157, 2018.

Zhang, X., Liu, L., Chen, X., Gao, Y., Xie, S., and Mi, J.: GLC_FCS30: Global land-cover product with fine classification system at 30m using time-series Landsat imagery, Earth System Science Data, 13, 2753–2776, https://doi.org/10.5194/essd-13-2753-2021, 2021.

Zhao, J., Ouyang, Z., Zheng, H., Zhou, W., Wang, X., Xu, W., and Ni, Y.: Plant species composition in green spaces within the built-up areas of Beijing, China, Plant ecology, 209, 189–204, https://doi.org/https://doi.org/10.1007/s11258-009-9675-3, 2010.

Zhou, W., Wang, J., Qian, Y., Pickett, S. T., Li, W., and Han, L.: The rapid but "invisible" changes in urban greenspace: A comparative study of nine Chinese cities, Science of the Total Environment, 627, 1572–1584, https://doi.org/https://doi.org/10.1016/j.scitotenv.2018.01.335, 2018.

Zhou, X. and Wang, Y.-C.: Spatial-temporal dynamics of urban green space in response to rapid urbanization and greening policies, Landscape and urban planning, 100, 268–277, https://doi.org/https://doi.org/10.1016/j.landurbplan.2010.12.013, 2011.

https://doi.org/10.5194/essd-2022-75
Preprint. Discussion started: 22 June 2022

Open Access Earth System Science Data Discussions

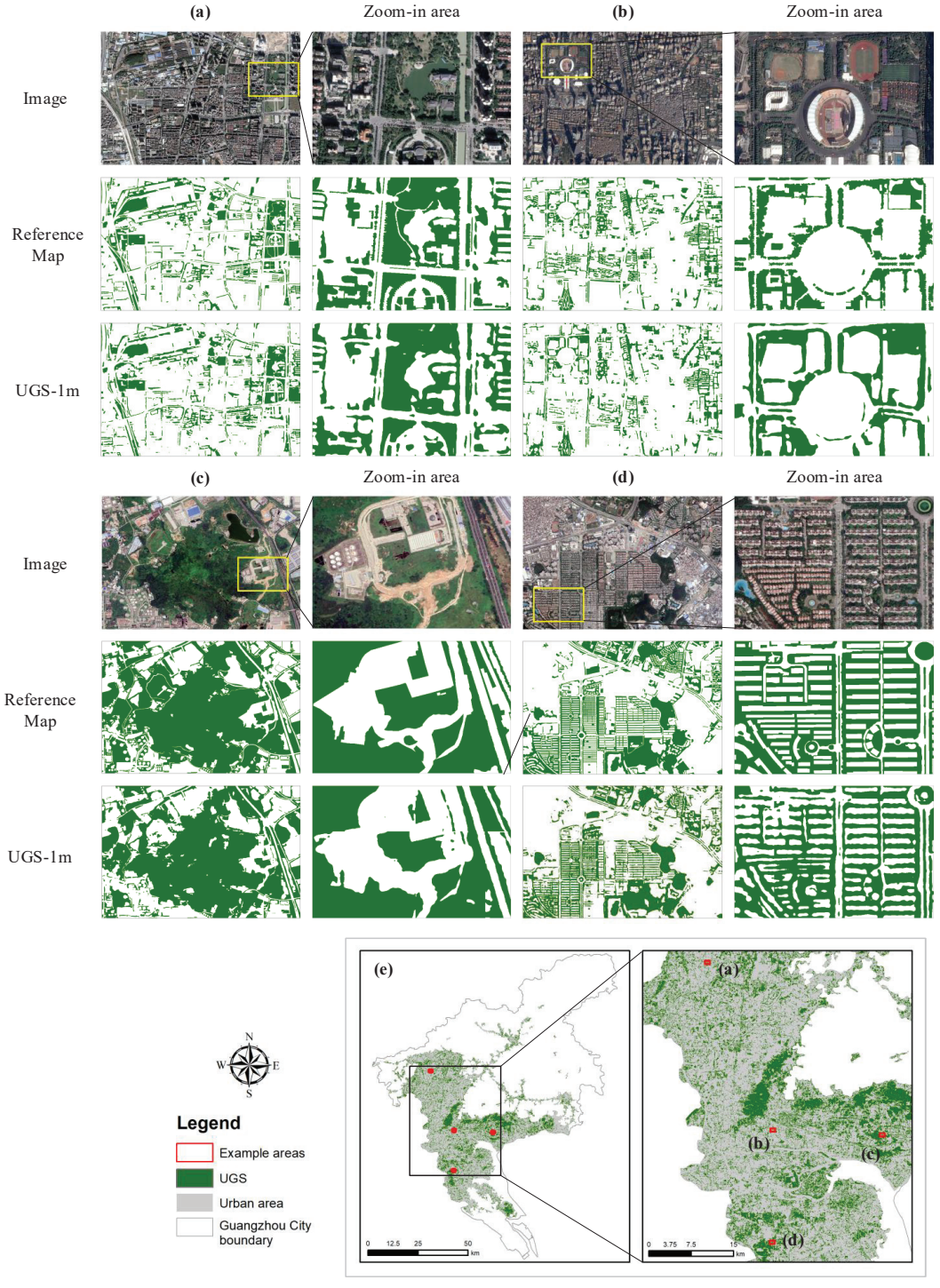


Figure 10. Visual comparisons of UGS results in UGS-1m: case study in Guangzhou City. (a)-(d) is four example areas collected from Guangzhou (Images © Google Earth 2020), (e) denoted the geographical location of (a)-(d).

https://doi.org/10.5194/essd-2022-75
Preprint. Discussion started: 22 June 2022

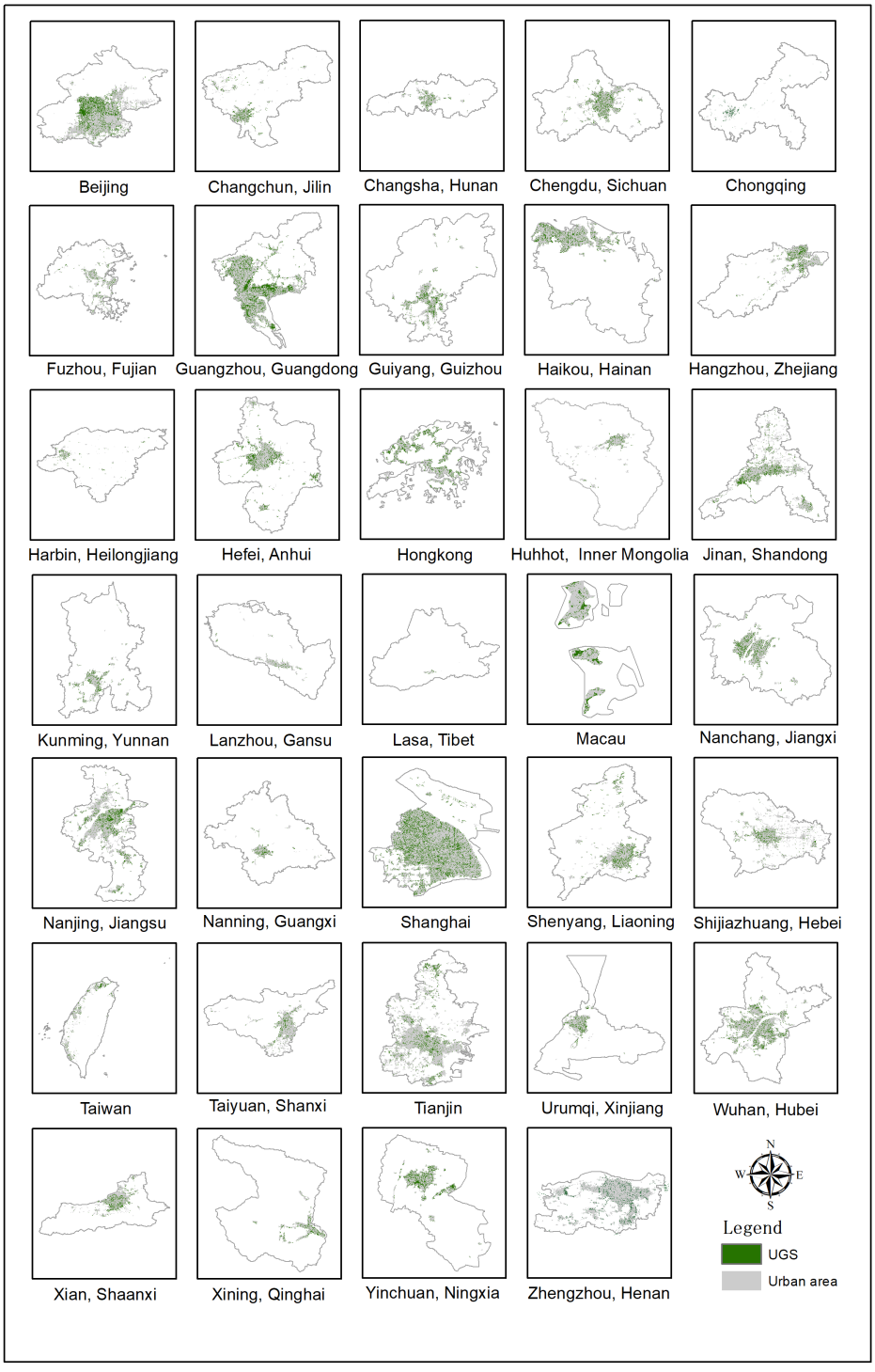


Figure 11. UGS results of the 34 major cities/areas in China (UGS-1m).

https://doi.org/10.5194/essd-2022-75
Preprint. Discussion started: 22 June 2022

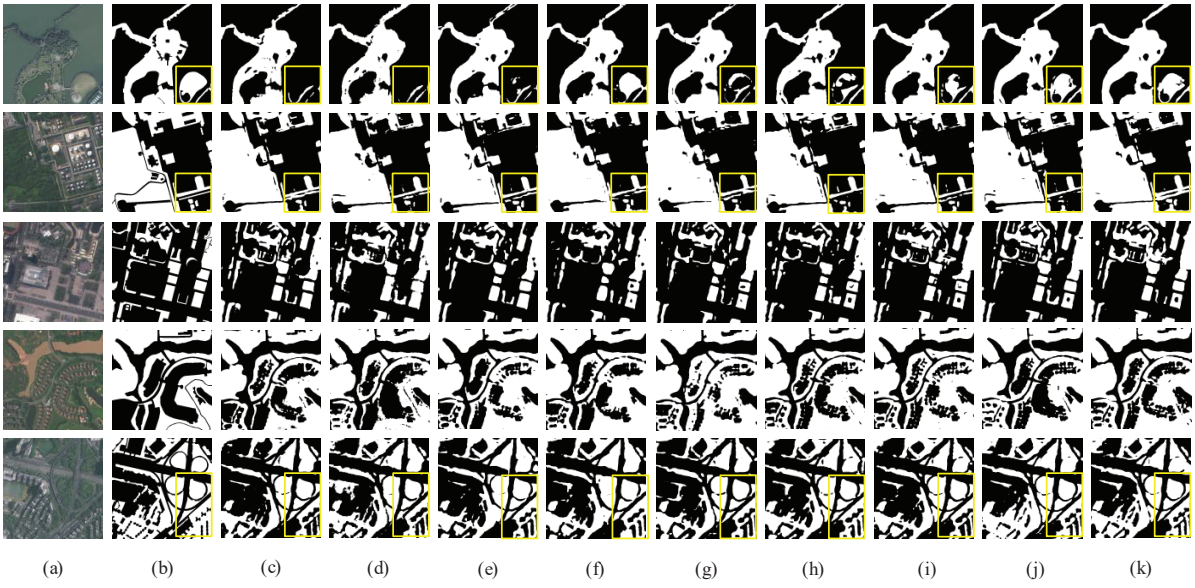


Figure 12. Visualization comparisons of different methods on UGSet (Images were retrieved from GF-2 2019). (a) Image; (b) Label; (c) SegNet; (d) UNet; (e) UperNet; (f) BiSeNet; (g) PSPNet; (h) Base; (i) Base+ECA; (j) Base+Point Head; (k) UGSNet.

https://doi.org/10.5194/essd-2022-75
Preprint. Discussion started: 22 June 2022

Open Access Earth System Science Data Discussions

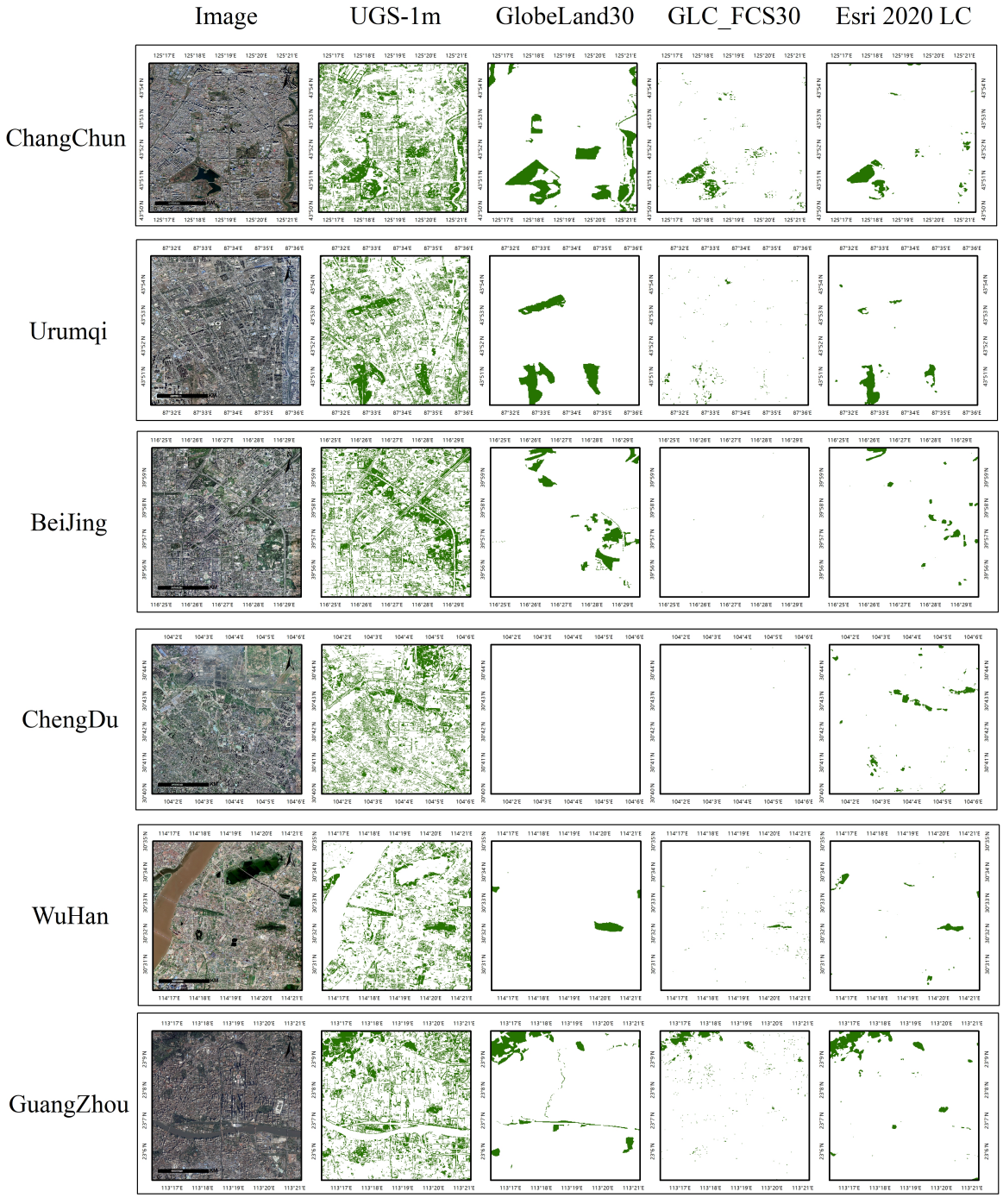


Figure 13. Visualization comparisons between UGS-1m, GlobeLand30 (Chen and Chen, 2018), GLC_FCS30 (Zhang et al., 2021) and Esri 2020 LC (© 2021 Esri) (Images © Google Earth 2020).