Land Use Land Cover Labeling of GLOBE Images Using a Deep Learning Fusion Model

Sergio Manzanarez, Vidya Manian * and Marvin Santos

Department of Electrical and Computer Engineering, University of Puerto Rico, Mayaguez, PR 00681, USA

* Correspondence: vidya.manian@upr.edu

Abstract: Most of the land use land cover classification methods presented in the literature have been conducted using satellite remote sensing images. High-resolution aerial imagery is now being used for land cover classification. The Global Learning and Observations to Benefit the Environment land cover image database, created by citizen scientists worldwide who use their handheld cameras to take a set of six images per land cover site. These images have clutter due to man-made objects, and the pixel uncertainties result in incorrect labels. The problem of accurate labeling of these land cover images is addressed. An integrated architecture that combines Unet and DeepLabV3 for initial segmentation, followed by a weighted fusion model that combines the segmentation labels, is presented. The land cover images with labels are used for training the deep learning models. The fusion model combines the labels of five images taken from the north, south, east, west, and down directions to assign a unique label to the image sets. 2916 GLOBE images have been labeled with land cover classes using the integrated model with minimal human-in-the-loop annotation. The validation step shows that our architecture of labeling the images results in 90.97% label accuracy. Our fusion model can be used for labeling large databases of land cover classes from RGB images.

Keywords: land use land cover classification; image segmentation; fusion model; GLOBE database; deep learning; DeepLab; UNet

1. Introduction

Land cover refers to the biological material on the surface of the Earth; on the other hand, it describes how people utilize the land and socio-economic activity. It also shows how people use the landscape for development, agricultural, or mixed uses. Thus, the Land Use Land Cover (LULC) refers to classifying the human activities or elements covering the Earth using scientific and statistical methods.

LULC classification methods are broadly categorized as pixel-based or object-based methods. Traditionally, the LULC was conducted manually, later moved to digital technology using the GIS technology, and recently, the Artificial Intelligence (AI) revolution has led to the emergence of new algorithms. A review of machine learning methods using feature extraction for supervised and unsupervised LULC classification is given in [1]. The AI methods for classification follow hard classification performed by traditional Machine Learning (ML) methods, which can be supervised, unsupervised, or semi-supervised; the second approach is soft classification. Recently, Deep Learning (DL) architectures are being widely used with satellite remote sensing images. High level feature extraction using the Deep Boltzman Machine (DBM) and low-level feature extraction using Principal Component Analysis (PCA) and Random Forest (RF) are compared in [2] for land cover classification from multispectral Lidar data. A hierarchical clustering method is Bag of Words (BoW), which is used to represent unstructured data as words. In [3], the BoW and Probabilistic Topic Model are used for extracting latent features for scene classification. The conventional neural networks with a few layers such as AlexNet, CaffeNet, and VGGNet have depth limitations that cannot gather semantic information, making them...
unsuitable for LULC classification. Moreover, a particular deep network is not versatile enough to tackle all LULC labeling problems. Hence, AlexNet, Inception-v3, and ResNet18 are combined in [4] with a new fully connected layer and a softmax layer to generate the LC labels. A transfer learning approach is presented in [5], where a pretrained network is used both in the feature descriptor and classifier stages. This approach is applied to aerial orthophotographs. A comparative study of transfer learning-based deep networks are given in [6]. The model performance is improved using enhancement techniques such as early stopping, gradient clipping, adaptive learning rates, and data augmentation. The model is applied for the LULC classification of satellite imagery.

The same scene in an image of a land cover may have complex spatial and structural patterns. In contrast, different land-use types may have a similar reflectance spectrum and texture. It becomes challenging to obtain a precise classification of the land use type due to high intraclass heterogeneity, and low interclass diversity [4]. We observe from the above that LULC classification is conducted on satellite imagery, and high resolution aerial photographs. Whereas, the GLOBE database is composed of land cover images taken by handheld cameras and have a higher resolution than a satellite or aerial image. Hence, we propose a segmentation pipeline that is novel in the following ways for labeling of these images:

1. The images are preprocessed using a pretrained DeepLabV3 architecture for removing man-made objects with a mask. Four architectures: DeepLabV3, DeepLabV3+, UNet, and UNet+ are used for segmentation.
2. The segmented directional images are combined in a decision block using a weighted fusion method to generate a label for the most common land cover class of the group of images.

In this paper, we present an integrated deep learning model for segmenting directional LULC images, combining the segmented output for generating unique labels to groups of images. The rest of the paper is organized as follows. Section 2 presents the preprocessing methods applied to the GLOBE images, architecture of the integrated UNet, UNet++, DeepLabV3, DeepLabV3+ models, and final labeling stage. Section 3 presents the results, while Section 4 discusses the results and compares them with those of the state-of-the-art methods. The conclusions are provided in Section 5.

2. Materials and Methods

This paper uses CNN-based UNet, Unet++, DeepLabV3, and DeepLabV3+ for semantic segmentation on the GLOBE Dataset.

2.1. GLOBE Image Database

The Global Learning and Observations to Benefit the Environment (GLOBE) application allows the public who are called citizen scientists to make observations of the environment using their camera and post the images in the GLOBE database. It is an open dataset available to researchers, students, and the scientific community for performing their investigation on different domains, such as the atmosphere, biosphere, hydrosphere, and pedosphere. The visualization tool [7] includes biometry with tree height data, land cover photos, lilac phenology, and carbon cycle with stored carbon and tree diversity. The atmosphere domain has data on air temperature, aerosols, clouds, precipitation, humidity, surface ozone, surface temperatures, and water vapor. GLOBE Observer (GO) land cover data collection began in September 2018. The observers voluntarily assign the land cover types to the images and are not accurate. The observers use the GO app to upload their photos. The GLOBE land cover data aims to verify remote sensing land cover classifications. The GO summary data consists of the following information: Land cover id, data source, measured at location, overall land cover classification, field notes, description of the six photos taken at the north, south, east, west, up, and down directions.

The series of six photos taken by observers are quality checked. The citizen scientists who capture the photos are kept anonymous. Many applications of the GLOBE land cover
data are listed in [8]. The photos in Figure 1 are taken at latitude 35.47° and longitude −83.32°. The volunteers do not always label the image with the land cover type, and many labels are erroneously given, as they do not have experience in land cover classification. However, this is an endeavor that is beneficial for many applications such as urban planning [9,10] and environmental protection [11].

![Sample images from GLOBE Visualization tool. (a) North, (b) South, (c) East, (d) West, (e) Up, (f) Down.](image)

### Table 1. Data augmentation.

| Augmentation | Unet | Unet+ | DeepLabv3 | DeepLabv3+ |
|--------------|------|-------|-----------|------------|
| Crop         | (450,512) | (450,512) | (450,512) | (450,512) |
| Rotation     | 90°, 180° and 270° | 90°, 180° and 270° | 90°, 180° and 270° | 90°, 180° and 270° |
| Brightness and Contrast | 0 to 0.05 | 0 to 0.05 | 0 to 0.05 | 0 to 0.05 |
| Flip x axis  | yes | yes | yes | yes |
| Flip y axis  | yes | yes | yes | yes |
| Median Blur  | 3 | 3 | 3 | 3 |
Annotations and Classes

The GLOBE database does not have ground truth. Hence, we annotated the images and created a ground truth for training and validation. The trained model can now be used without human in the loop for labeling new images. The images are annotated using 12 classes: Grass, Building, Tree, Road, Sky, Soil, Bare Rock, Sand, Sidewalk, Water, Gravel, Blue Mountain as shown in Table 2. Annotations were made manually with Labelme [13] using polygon shape as shown in Figure 2.

Table 2. Semantic class used for train segmentation models stage.

| Class ID | Name           | Description                                                                 |
|----------|----------------|-----------------------------------------------------------------------------|
| 1        | Grass          | High grass, low grass, grassland and herbaceous.                             |
| 2        | Building       | Building, houses and bridges.                                                |
| 3        | Tree           | Trees or shrubs, that are part of a forest, include stem and main wooden axis of a tree and the tree crown. |
| 4        | Roads          | Surfaces that are made of asphalt used for vehicular traffic, including parking lots, but only those made of asphalt. |
| 5        | Sky            | Open sky, cloud, without leaves of trees or any other occlusions             |
| 6        | Soil           | Surfaces that are made of dry or wet soil e.g., soil, dirt roads, muddy or bare soil. |
| 7        | Bare Rock      | Bare rock places, including rocks.                                           |
| 8        | Sand           | Beach sand or river sand.                                                    |
| 9        | Sidewalk       | Pedestrian traffic places that are made of concrete.                         |
| 10       | Water          | Flowing and stagnant water.                                                  |
| 11       | Gravel         | Medium or small gravel used for construction or filling of pedestrian or vehicular paths. |
| 12       | Blue Mountain  | Only mountains look like blue due to the far viewing distance. Area not classified to any class. It can include e.g., man-made objects, persons, cars, and all objects excluded from above. In general all unlabeled pixels are defined with the label “undefined”. |
| 0        | Background     |                                                                              |

Figure 2. Example of manual labeling image of GLOBE dataset. (a) Original Image, (b) image labeling.

2.3. CNN for Semantic Segmentation

We work with CNN-based models due to their high performance in high-resolution images for per-pixel semantic labeling (or image classification) [14]. Specifically, this paper focuses on semantic segmentation using deep learning architectures, as it has more advantages than traditional image processing techniques [15,16]. Semantic segmentation deep learning models can generally be viewed as an encoder network followed by a decoder network. An encoder maps the input into a code, usually a pre-trained classification network such as VGG/Resnet, and a decoder maps the code to reconstruct the input. In semantic segmentation of LULC images, each pixel is given a unique label. One of the first Deep Convolutional Neural Networks (DCNNs) used for semantic segmentation is...
the Fully Convolutional Network (FCN) [17]. The FCN network model is an extension of the classical CNN but is a network consisting of only convolutional layers. Various more advanced FCN-based approaches have been proposed, such as U-Net, DeepLab, DeeplabV3, U-Net++, SegNet, and DeeplabV3+. This paper focuses on four deep learning semantic segmentation architectures: U-Net, DeepLabV3, U-Net++, and DeepLabV3+. U-Net is an image segmentation model proposed by Ronneberger et al. [18], which is characterized by symmetrical U-shape architecture consisting of the symmetric contracting path and expansive path, as shown in Figure 3a. U-Net modifies and extends the FCN; the main idea is to make FCN maintain the high-level features in the early layer of the decoder. To this end, they use concatenation for fusing decoder blocks by skipping connections to localize the segmentations. Similarly, Unet++, introduced by Zhou et al. in [19], is an architecture for semantic segmentation based on both nested and dense skip connections. The decoder of Unet++ is more complex than in Unet. In Figure 3b, black blocks refer to the original U-Net, while the green show dense convolution blocks on the skip pathways.

DeepLabV3 is a semantic segmentation model designed and open-sourced by Google that outperforms DeepLabV1 and DeepLabV2. In DeepLab models, input feature maps become smaller while traversing through the convolutional and pooling layers of the network by using Atrous convolutions and Atrous Spatial Pyramid Pooling (ASPP) modules [20,22]. DeepLab-v3+ is an extended model based on the DeepLab-v3 model by adding a decoder module to refine the segmentation results, especially along object boundaries. The depth-wise separable convolution is applied to Atrous spatial pyramid pooling and decoder modules, resulting in a faster and stronger encoder-decoder network for semantic segmentation [21].
2.4. Workflow for Supervised Classification

Figure 4 illustrates the proposed labeling procedure workflow. The workflow has five stages; ① the first is the preprocessing stage, which corresponds to our annotated datasets, and then data augmentation. ② After this, four deep learning semantic segmentation models were trained, UNet, UNet++, DeepLabV3, and DeepLabV3+; implementation of these models follows [23]. ③ We use DeepLabV3 [20] architecture to remove some man-made objects that appear in GLOBE images. In general, the pre-trained DeepLabV3 with Resnet50 as a backbone model is used to detect man-made objects in the third stage; for example, most of the down direction images have the feet of the photographer, which are removed using DeepLabV3. This pre-trained model has been trained on a subset of Common Objects in the Context dataset (COCO dataset) [24], on the 20 categories present in the Pascal VOC dataset, and was implemented on PyTorch [25]. Those 20 categories on Pascal VOC are the aeroplane, bicycle, bird, boat, bottle, bus, car, cat, chair, cow, diningtable, dog, horse, motorbike, person, pottedplant, sheep, sofa, train, and tvmonitor. ④ In the fourth step, we remove these man-made objects by pointwise multiplication of the mask generated in step three and each segmented image generated by each semantic segmentation model.

Figure 4. Block diagram for the land cover classification framework. The arrows represent the input and output for each stage, and the stage are represented by the blocks.

⑤ The last step is the final fusion step that combines the segmentation results of the four models. The postprocessing step builds a new classifier by taking a majority vote of the four segmentation models in the previous step. If we denote the segmentations models as $M_i$ for $i \in \{1, 2, 3, 4\}$. For each pixel $x$, each model $M_i$ generates a probability vector $P_i(x) = (p_{i,1}(x), p_{i,2}(x), \ldots, p_{i,13}(x))$, where $p_{i,j}(x)$ represent the probability of the pixel $x$ belonging to class $j$. These models predict class label for $x$ as $C_i(x) = \arg\max_{j=1,2,\ldots,13} P_i(x))$. Next, the final segmentation model predicts a class label for each $x$ as $C(x) = \text{mode}\{C_1(x), C_2(x), C_3(x), C_4(x)\}$. If the mode is not unique, we randomly select one of the four models. Finally, the percentage of LC label from the five directional images is divided by the number of pixels of each class and the total number of pixels of the directional images. We do not use the Up image from GLOBE (see Figure 1) because it does not improve the information for land cover classes, in most images there are only sky and/or clouds. We also consider the perspective of the five images taken by weighting the down directional image. If $p_j$
represents the percent of land cover class \( j \) in the directional image and if \( I_1, I_2, I_3, I_4 \) and \( I_5 \) represents the North, South, East, West, and Down images, respectively, then

\[
p_j = \frac{\sum_{k=1}^{5} \sum_{x \in I_k, C(I_k(x)) = j} w_k}{\sum_{k=1}^{5} \sum_{x \in I_k} w_k}
\]

2.5. Supervised Semantic Segmentation

For training semantic segmentation models, we use the Adam optimizer with a learning rate of 0.001. The cost function is essential in the training and validation stages. However, there are many of them, and it is impossible to choose one of them as the best [26] for the segmentation task. As shown in Figure 5, our dataset is highly unbalanced. Hence, the Dice Loss function [27] is selected for its high performance.

\[
\text{Dice Coeff} = \frac{2y \cap y_{\text{pred}}}{y + y_{\text{pred}}}, \quad \text{Dice Loss} = 1 - \text{Dice Coeff}
\]

All our models are trained for 1000 epochs, with a batch size of seven. The validation metrics are the mean intersection over unit and labeling accuracy. Training the four models with 600 images takes four days in Google Colab.

Figure 5. Histogram of the class distribution on the training dataset. The \( y \)-axis is in logarithmic scale.

A total of 600 images are used for training the networks, and 100 images for the validation for 13 land cover classes. The output labels are verified for accuracy with testing images from existing correctly labeled GLOBE images. The integrated model is used to assign LC labels to 2915 GLOBE images. Supplementary Table S1 gives the GPS positions and the links to download the five directional images for the database of 2916 LC images. The Supplementary Table S2 gives the predicted labels for the 2916 groups of five image database, totaling the fusion method’s use of 14,580 individual images. Four labels are assigned to each of the 2915 group of five images ordered with higher to lower probability of the LC class found in the image as Label 1, Label 2, Label 3, and Label 4. Land cover label 1 is the most encountered LC class in the five directional images, followed by the other three labels.

In order to standardize the image acquisition and sharing of LULC images around the world, certain requirements can be imposed on the citizen scientists for acquisition and sharing of LULC images such as (a) use of certain models of cameras, (b) imaging distance, (c) imaging time during the day, and (d) weather conditions such as cloud cover during the day, so that the illumination is uniform for all the acquired images.
3. Results

This section presents and discusses the results of applying the proposed method for labeling and classifying the GLOBE images. To evaluate and analyze the performance of UNet, UNet++, DeepLabV3, DeepLabV3+, and the fusion method proposed, we use Intersection over Union (IoU), also known as the Jaccard index shown in Figure 6. This metric is defined as the similarity of two regions based on their overlap.

![Image of IoU](image_url)

**Figure 6.** Intersection over union.

In Table 3, we list and compare the segmentation results obtained on the validation set using the IOU metric for UNet, UNet++, DeepLabV3, DeepLabV3+, and the proposed method. The results demonstrate that our method improves the IOU. The challenging classes such as Building, Sidewalk, and Road also have good accuracies.

| Class       | UNet  | UNet++ | DeepLabV3 | DeepLabV3+ | Fusion |
|-------------|-------|--------|-----------|-------------|--------|
| Grass       | 0.8496| 0.9018 | 0.9083    | 0.8842      | 0.9102 |
| Building    | 0.7344| 0.7400 | 0.6390    | 0.7315      | 0.7656 |
| Tree        | 0.7825| 0.8148 | 0.8712    | **0.8906**  | 0.8787 |
| Roads       | 0.7668| 0.8611 | 0.8592    | 0.8277      | 0.8905 |
| Sky         | 0.9086| 0.9802 | 0.9623    | 0.9721      | 0.9791 |
| Soil        | 0.7947| 0.8280 | 0.7526    | 0.8067      | 0.9092 |
| Bare Rock   | 0.9576| 0.9247 | 0.8653    | 0.9735      | 0.9813 |
| Sand        | 0.7926| 0.8221 | 0.9968    | 0.9967      | 0.9891 |
| Sidewalk    | 0.7466| 0.8903 | 0.8795    | **0.8970**  | 0.8966 |
| Water       | 0.9349| 0.8605 | 0.8665    | 0.9330      | 0.9413 |
| Gravel      | 0.7563| 0.5810 | 0.4314    | 0.9494      | 0.9350 |
| Blue Mountain| 0.6635| 0.9285 | 0.8206    | 0.8249      | 0.8394 |
| Background  | 0.8335| 0.8814 | 0.6960    | 0.7799      | 0.8026 |

The IOU of our fusion model increased accuracy by 1.91% on average compared to the best of the four individual models. However, if we look at the soil class, the IOU increased by 2.94%. This result demonstrates that our method performs better than the existing models for segmenting directional, rotated and translated land cover imagery.

For the image sets in Figure 7, the land cover class labels obtained with our fusion model is given in Table 4.

The confusion matrix with land cover labeling accuracies for the 13 classes using the fusion model is given in Figure 8. The classification accuracy is calculated as the number of pixels correctly classified over the total number of pixels for each land cover class multiplied by 100. The model obtained an Overall Accuracy (OA) of 90.97%. The black diagonal entries have accuracies closer to 100%. The off-diagonal entries have lower accuracies indicating better performance of the fusion model.
Figure 7. Three sets of five directional GLOBE images and their respective segmented images with the fusion model.
Table 4. Land cover and location of the two directional images.

|       | Bare Rock 33.20% | Gravel 22.40% | Sky 19.06% | Sand 18.99% | Latitude 54.4206 | Longitude −0.523 |
|-------|------------------|---------------|------------|-------------|------------------|------------------|
| Img1  |                  |               |            |             |                  |                  |
|        |                   |               |            |             |                  |                  |
| Img2  | Grass 49.43%     | Roads 24.04%  | Sky 11.58% | Tree 9.25%   | Latitude 44.5668 | Longitude −123.2958 |
|        |                   |               |            |             |                  |                  |
| Img3  | Grass 45.84%     | Sky 22.75%    | Background 18.57% | Tree 8.40%   | Latitude 42.0859 | Longitude −83.2099 |

4. Discussion

The GLOBE images are obtained from handheld cameras; hence, they have a lot of variability in each land cover class. The proposed fusion model has a validation mIOU of 90.97%. The method has difficulty discriminating between the buildings, road, sidewalk classes, and grass and tree classes.

Figure 9 shows the difference in the mean intensity values for each class from images taken from different GPS locations. It can be observed that the building class has higher variability, and hence the accuracy in the confusion matrix is lower for this class.

The images are complex, as they are high-resolution RGB images taken from urban and rural areas with a high amount of detail and man-made objects. The images need a lot of preprocessing to remove the noise and variability due to multiple classes and be assigned a land cover label. Our fusion architecture is scalable and can be expanded to a larger application to include more land cover classes. We have developed the fusion architecture in GoogleColab, which runs the four models in parallel using the Graphical Processing Unit (GPU). The current architecture uses four models, but the fusion can be conducted with three models. At least three models are necessary for majority voting. We obtained good results by fusing four models.
Each of the architectures perform well for a particular land cover class, and not for the other. For example, Unet performs well for water and is useful for removing background objects, and DeepLabV3+ performs well for grass and sand classes. Hence, it is necessary to check the accuracies individually for each land cover class.

5. Conclusions

A fusion method combining four deep learning architectures is used for labeling high resolution remote sensing imagery from the GLOBE database. DeepLabV3+ is used to remove man-made objects from the images. A weighted averaging is used in the decision block to combine the segmentation results of the four architectures for generating land cover labels with probabilities. The accuracy of labeling achieved is 90.97% for the validation dataset. The GLOBE database does not have ground truth, and our proposed method uses minimal human in the loop based annotation. Our method successfully labeled 2916 image groups. A total of 20% of the images are used for training, and 7% for validation. This work provides a tool for the labeling of land cover images acquired by the public from different parts of the world with different camera settings as well as random image captures from humans with no fixed rotation, translation, or scaling parameters. Despite the variability of the images in the GLOBE database, our labeling tool assigns labels with minimal human in the loop annotation.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/s22186895/s1, Table S1: 2915 GLOBE images with GPS positions and links for downloading the five directional images. Table S2: Predicted land cover labels using the fusion method. Labels are ordered from higher to lower probability.

Author Contributions: Conceptualization, V.M. and S.M.; methodology V.M., S.M. and M.S.; software, S.M.; validation, S.M. and M.S.; formal analysis V.M., S.M. and M.S.; investigation, V.M., S.M. and M.S.; data curation, S.M. and M.S.; writing—original draft preparation, V.M., S.M. and M.S.; writing—review and editing, V.M., S.M. and M.S.; visualization, V.M., S.M. and M.S.; supervision,
V.M.; project administration, V.M.; funding acquisition, V.M. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by NASA EPSCoR, grant number 80NSSC21M0156. Opinions, findings, conclusions, or recommendations expressed in this material are those authors and do not reflect the views of NASA.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The Biosphere Land Cover Classification image datasets are available at [https://vis.globe.gov/GLOBE/](https://vis.globe.gov/GLOBE/) (accessed on 31 March 2022).

**Acknowledgments:** The material contained in this document is based upon work supported by National Aeronautics and Space Administration (NASA) grant 80NSSC21M0156. Thanks to Estefanía Alfaro-Mejía for her valuable comments.

**Conflicts of Interest:** The authors declare no conflict of interest.

**References**

1. Alshari, E.A.; Gawali, B.W. Development of classification system for LULC using remote sensing and GIS. *Glob. Transit. Proc.* 2021, 2, 8–17. [CrossRef]
2. Pan, S.; Guan, H.; Yu, Y.; Li, J.; Peng, D. A Comparative Land-Cover Classification Feature Study of Learning Algorithms: DBM, PCA, and RF Using Multispectral LiDAR Data. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 2019, 12, 1314–1326. [CrossRef]
3. Shao, H.; Li, Y.; Ding, Y.; Zhuang, Q.; Chen, Y. Land Use Classification Using High-Resolution Remote Sensing Images Based on Structural Topic Model. *IEEE Access* 2020, 8, 215943–215955. [CrossRef]
4. Xu, L.; Chen, Y.; Pan, J.; Gao, A. Multi-Structure Joint Decision-Making Approach for Land Use Classification of High-Resolution Remote Sensing Images Based on CNNs. *IEEE Access* 2020, 8, 42848–42863. [CrossRef]
5. Zhao, B.; Huang, B.; Zhong, Y. Transfer Learning with Fully Pretrained Deep Convolution Networks for Land-Use Classification. *IEEE Geosci. Remote Sens. Lett.* 2017, 14, 1436–1440. [CrossRef]
6. Naushad, R.; Kaur, T.; Ghaderpour, E. Deep Transfer Learning for Land Use and Land Cover Classification: A Comparative Study. *Sensors* 2021, 21, 8083. [CrossRef] [PubMed]
7. The Globe Program. Available online: [https://vis.globe.gov/GLOBE/](https://vis.globe.gov/GLOBE/) (accessed on 5 January 2022).
8. Kohl, H.A.; Nelson, P.V.; Pring, J.; Weaver, K.L.; Wiley, D.M.; Danielson, A.B.; Cooper, R.M.; Mortimer, H.; Overoye, D.; Burdick, A.; et al. GLOBE Observer and the GO on a Trail Data Challenge: A Citizen Science Approach to Generating a Global Land Cover Land Use Reference Dataset. *Front. Clim.* 2021, 3, 620497. [CrossRef]
9. Pauleit, S.; Duhme, F. Assessing the environmental performance of land cover types for urban planning. *Landsc. Urban Plan.* 2000, 52, 1–20. [CrossRef]
10. Zhou, W.; Huang, G.; Cadenasso, M.L. Does spatial configuration matter? Understanding the effects of land cover pattern on land surface temperature in urban landscapes. *Landsc. Urban Plan.* 2011, 102, 54–63. [CrossRef]
11. Pereira, H.M.; David Cooper, H. Towards the global monitoring of biodiversity change. *Trends Ecol. Evol.* 2006, 21, 123–129. [CrossRef] [PubMed]
12. Buslaev, A.; Iglovikov, V.I.; Khvedchenya, E.; Parinov, A.; Druzhinin, M.; Kalinin, A.A. Albumentations: Fast and flexible image augmentations. *Information* 2020, 11, 125. [CrossRef]
13. Wada, K.; mpitid ; Buijs, M.; Zhang, C.N.; Kubovčík, B.M.; Myczko, A.; latenti; Zhu, L.; Yamaguchi, N.; Fujii, S.; et al. Labelme: Image Polygonal Annotation with Python. 2021. Available online: [https://zenodo.org/record/5711226#.YxL6CHZBxPY](https://zenodo.org/record/5711226#.YxL6CHZBxPY) (accessed on 15 June 2022).
14. Zhang, P.; Ke, Y.; Zhang, Z.; Wang, M.; Li, P.; Zhang, S. Urban Land Use and Land Cover Classification Using Novel Deep Learning Models Based on High Spatial Resolution Satellite Imagery. *Sensors* 2018, 18, 3717. [CrossRef] [PubMed]
15. Ali, M.B.; Bai, X.; Gu, I.Y.H.; Berger, M.S.; Jakola, A.S. A Feasibility Study on Deep Learning Based Brain Tumor Segmentation Using 2D Ellipse Box Areas. *Sensors* 2022, 22, 5292. [CrossRef] [PubMed]
16. Lin, F.; Gan, L.; Jin, Q.; You, A.; Hua, L. Water Quality Measurement and Modelling Based on Deep Learning Techniques: Case Study for the Parameter of Secchi Disk. *Sensors* 2022, 22, 5399. [CrossRef] [PubMed]
17. Long, J.; Shelhamer, E.; Darrell, T. Fully Convolutional Networks for Semantic Segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, USA, 7–12 June 2015.
18. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Proceedings of the Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015*, Munich, Germany, 5–9 October 2015; Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F., Eds.; Springer International Publishing: Cham, Switzerland, 2015; pp. 234–241.
19. Zhou, Z.; Rahman Siddiquee, M.M.; Tajbakhsh, N.; Liang, J. UNet++: A Nested U-Net Architecture for Medical Image Segmentation. In Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support; Stoyanov, D., Taylor, Z., Carneiro, G., Syeda-Mahmood, T., Martel, A., Maier-Hein, L., Tavares, J.M.R., Bradley, A., Papa, J.P., Belagiannis, V., et al., Eds.; Springer International Publishing: Cham, Switzerland, 2018; pp. 3–11.

20. Chen, L.C.; Papandreou, G.; Schroff, F.; Adam, H. Rethinking atrous convolution for semantic image segmentation. arXiv 2017, arXiv:1706.05587.

21. Chen, L.C.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. Encoder-decoder with atrous separable convolution for semantic image segmentation. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018, pp. 801–818.

22. Chen, L.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. IEEE Trans. Pattern Anal. Mach. Intell. 2018, 40, 834–848. [CrossRef] [PubMed]

23. Yakubovskiy, P. Segmentation Models Pytorch. 2020. Available online: https://github.com/qubvel/segmentation_models.pytorch (accessed on 4 July 2022).

24. Lin, T.Y.; Maire, M.; Belongie, S.; Hays, J.; Perona, P.; Ramanan, D.; Dollár, P.; Zitnick, C.L. Microsoft COCO: Common Objects in Context. In Proceedings of the Computer Vision—ECCV 2014, Zurich, Switzerland, 6–12 Septwmber 2014; Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T., Eds.; Springer International Publishing: Cham, Switzerland, 2014; pp. 740–755.

25. Paszke, A.; Gross, S.; Chintala, S.; Chanan, G.; Yang, E.; DeVito, Z.; Lin, Z.; Desmaison, A.; Antiga, L.; Lerer, A. Automatic differentiation in PyTorch. In Proceedings of the NIPS-W, Long Beach, CA, USA, 4–9 December 2017.

26. Jadon, S. A survey of loss functions for semantic segmentation. In Proceedings of the 2020 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB), Viña del Mar, Chile, 27–29 October 2020; pp. 1–7. [CrossRef]

27. Sudre, C.H.; Li, W.; Vercauteren, T.; Ourselin, S.; Jorge Cardoso, M. Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. In Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support; Springer: Berlin/Heidelberg, Germany, 2017; pp. 240–248.