Automatically Mapping Urban Green Space Using Sentinel-2 Imagery and Deep Learning Methods in Multiple Cities Worldwide: A Convolutional Neural Network Approach

Thesis project for Master of Applied Data Science in Utrecht University

July 2022

Student name	Jiawei Zhao
Student number	4768876
Student e-mail	j.zhao3@students.uu.nl
Supervisor	Dr. Labib S.M.
Supervisor e-mail	s.m.labib@uu.nl

Abstract

Urban green spaces (UGSs) are significant to the urban ecosystem and can have a positive influence on human health, both physically and mentally. Traditional methods of mapping UGS are mainly field surveys, however, the development of (very) high-resolution satellite images, together with machine learning and deep learning techniques has provided a faster way of extracting UGS at a large scale. This paper employed convolutional neural networks (CNNs) to conduct semantic segmentation of UGS from Sentinel-2 imagery for multiple cities around the globe. The main architectures used were U-Net model from base level and U-Net model with ResNet-50 and VGG-16 backbones pretrained on ImageNet. Results show that U-Net with pretrained ResNet-50 backbone achieved the best performance, with the average OA, IoU, F-score, and AUC of 0.9695, 0.9188, 0.9572, and 0.9584. When testing our model on external cities, the metrics above became 0.8767, 0.5570, 0.6448, and 0.6534, respectively, showing a relatively good generalization ability of the model to predict UGSs where benchmark city data is not available, for instance, cities in the global south. Moreover, our model identified some UGSs which are not in the ground truth datasets because of the hard definition of UGS, such as golf courses and cemeteries. However, since these areas indeed provide similar functions and benefits as UGSs, we argue that our model might be a potentially effective way to get more complete datasets for UGS distributions.

Keywords: urban green space, convolutional neural network, satellite imagery, semantic segmentation, transfer learning

Table of Contents

1. Introduction

Urban green spaces (UGSs), defined as accessible urban land covered by vegetations (World Health Organization, 2017), are a significant part of the urban ecosystem and urban aesthetics. UGSs can provide shelters for nature species, thus promoting urban biodiversity (Aronson et al., 2017). Moreover, due to the shadow vegetations create, sun radiation they deflect, and moisture they release into the atmosphere, UGSs can generate a cooling effect and mitigate the urban heat island effect in big cities (Yu & Hien, 2006). The ecological benefits UGSs bring above can ultimately have a positive influence on human health, both physically and mentally (Markevych et al., 2017; Twohig-Bennett & Jones, 2018). To be specific, UGSs provide a safe and healthy place for citizens to do recreational, social, and sports activities, reducing stress and loneliness. Additionally, UGS can reduce the threatening of high temperature, air pollution, and loud noise to the wellbeing of urban dwellers, decreasing the possibilities of cardiovascular, respiratory, and metabolic diseases (Aronson et al., 2017; Hartig et al., 2014; Markevych et al., 2017; Twohig-Bennett & Jones, 2018). Therefore, sustainable development goals (SDGs) have indicated that universal access to UGS should be provided by 2030 (SDG 11.7), especially for vulnerable groups[1] to make cities and human settlements inclusive, safe, resilient, and sustainable.

Various research has been conducted on modeling and mapping UGSs. One important part is to measure the accessibility of UGS to citizens (Wang et al., 2021). These studies usually analyzed urban population's accessibility to UGS considering distances as well as other criteria (e.g., size), thus revealing the equality or inequality of green space distribution. There is also research focused on measuring how green UGSs are using the Green Index calculated by the percentage area of green within an areal unit (Gupta et al., 2012). However, to accurately measure UGSs, understanding the distribution of them (i.e., where are UGSs and where are not) is the basic guarantee to better investigate how such spaces might be useful for improving the health of urban citizens.

Traditional inventories of UGS distribution datasets are mainly derived from field surveys (Chen et al., 2021) and visual interpretation of aerial imagery (Huerta et al., 2021). Such manual digitalization

[1] Women and children, older persons and persons with disabilities

methods are high in accuracy but intensive and require domain experts (Liu et al., 2019). In addition, city growth makes green spaces change over time (Sperandelli et al., 2013), leading to frequent updating needs of these datasets, which is costly and time-consuming. Meanwhile, these datasets are often unavailable in developing countries and regions, where the need for UGSs is more intensive (Albert et al., 2017; European Environment Agency, 2022; Shuvo et al., 2020).

With the availability of high-resolution satellite imageries and the development of machine learning and deep learning techniques, there are attempts to automatically create such a dataset. These methods show advantages over traditional ones due to the reduction of sources needed (Huerta et al., 2021) and coverage of larger areas. After training and testing the machine learning and deep learning models with remote sensing images, the models can learn the features of UGSs according to the input benchmark (the ground truth data for where are UGSs and where are not) and automatically identify UGS areas from satellite images for different time or locations. Many studies have been conducted on this topic these years (Albert et al., 2017; Huerta et al., 2021; Kranjčić et al., 2019; More et al., 2020; Pritt & Chern, 2020; Xu et al., 2020).

However, these attempts are mainly conducted at a local scale (e.g., one area or city), such as Inner Mongolia and Hubei Province in China (Liu et al., 2019), State of Nuevo Leon in the Mexican Republic (Huerta et al., 2021). As Albert et al. (2017) suggested, adding a diverse set of multiple regions or cities into the model can provide robust results. Moreover, the satellite imageries used in these studies were mostly Very High Resolution (VHR) ones, for example, WorldView-2/3 (Hartling et al., 2019) with a resolution of 0.46-1.8m (European Space Agency, 2022c). Nevertheless, VHR satellite images are not freely accessible, posing limitations on public research. So, we used Sentinel-2 as the satellite images repository, which is a freely accessible dataset with a resolution of 10-60m. Studies have shown that Sentinel-2 can be used for vegetation mapping (Chen et al., 2021), but due to its medium resolution, it fails to delineate fine-grained urban structures, namely trees, shrubs, and small buildings (Ludwig et al., 2021).

Based on the argument above, we aimed to automatically model and map UGS at multiple cities around the globe, pooling cities with various structures together. Also, we fused multi-source satellite and local datasets to see if better accuracy can be attained when using deep learning models for such activities (Ludwig et al., 2021). Additionally, our research put focuses on public UGSs since these are

the green areas accessible to the public. Therefore, our research question is: can we build a deep learning (e.g., Convolutional Neural Network – CNN) model to automatically identify public UGS from Sentinel-2 imagery in multiple cities worldwide? To answer this research question, several deep learning model architectures were tested, with different combinations of bands from satellite imageries and local data obtained from multiple data sources.

The rest of this paper is organized as follows: in section 2, we review the related research. In section 3, we elaborate on the data and methods used. The results of our models are shown in section 4, and discussion is in section 5. We conclude in section 6. All the codes used to process data, train, and evaluate models are available at https://github.com/mar-koz22/Park-NET-identifying-Urban-parks-using-multi-source-spatial-data-and-Geo-AI/tree/main/Jiawei's-approach.

2. Literature Review

In this section, three parts of related research are reviewed, including the development of satellite images, machine learning and deep learning techniques employed in UGS detection and mapping, and transfer learning methods used in such studies.

2.1 High-Resolution Satellite Images

The emergence of (very) high-resolution satellite imagery is the basic guarantee for automatically mapping of UGS. Compared with fine spatial resolution images (with a resolution between 5-100m), VHR images (with a resolution smaller than 5m) (Yao, 2009) can capture more details of land surface structures and distinguish even individual tree crowns and are thus widely used for urban tree species classification (Hartling et al., 2019), urban green space detection (Liu et al., 2019; Xu et al., 2020), and land use classification (Albert et al., 2017). Examples of VHR satellite image repositories include WorldView-2/3 (European Space Agency, 2022c), GeoEye-1 (European Space Agency, 2022a), Quickbird-2 (European Space Agency, 2022), and Ikonos (Satellite Imaging Corporation, 2022). Among them, WorldView-2/3 has a <3m multispectral resolution on eight spectral bands, while others only provide traditional four-spectral-band images (blue, green, red, and near-infrared). Apart from

VHR imageries, Landsat 8-9 (with a resolution of 15-100m) (U.S. Geological Survey, 2022) and Sentinel-2 (with a resolution of 10-60m) (European Space Agency, 2022b) are two freely accessible datasets with fine spatial resolution and are popularly used in UGS mapping (Labib & Harris, 2018).

Moreover, with the help of multispectral images, calculated as the ratios of satellite image bands, vegetation indices play an important role in vegetation analysis. Various vegetation indices have been proposed, and normalized difference vegetation index (NDVI) is one of the most successful attempts to simply and quickly identify vegetated areas (Earth Observatory, 2022), and thus is widely used in research. For example, Xu et al. (2020) calculated the difference between NDVI in summer and winter and added this as another layer in addition to the red, green, and blue bands retrieved from remote sensing images when doing the change detection of UGS in Beijing.

2.2 Machine Learning and Deep Learning Techniques

The advancement of computational technologies has become the other premise to make automatic UGS detection or land use classification possible using machine and deep learning methods. Support Vector Machine (SVM), naïve Bayes classifier, and random forest (RF) are the most often used machine learning algorithms when conducting classification or change detection studies. More et al. (2020) used an SVM classifier to detect UGS changes using spatio-temporal data from MODIS and Landsat 7 datasets. They found the overall green spaces in Mumbai metropolitan area in India reduced to 50% over 15 years. The SVM, RF, and naïve Bayes classifiers were employed by Kranjčić et al. (2019) to do green infrastructure classification in city areas, and they concluded that SVM outperformed the other two classifiers according to its classification accuracy and performance time.

However, the machine learning classifiers above use low-level features, such as spectral, textural, geometrical, and contextual features, and this may lead to extract and classification errors (Chen et al., 2021). Different from this, deep learning algorithms can extract high-level features (e.g., objects) and are proven to have a good performance in object detection, image classification, and semantic segmentation (Huerta et al., 2021). When applied to satellite images, the most popular deep learning approaches are semantic segmentation and classification, and these rely on the implementation of different Convolutional Neural Network (CNN) architectures. CNNs are efficient models which can extract high-level features using sets of convolution, activation, and pooling operations (LeCun et al.,

2015). However, CNNs themselves cannot do a per-pixel segmentation since they tend to gradually aggregate information over a larger and larger size (Liu et al., 2019). Due to the absence of a dense layer, Fully Convolution Networks (FCNs) manage the layers in such a way that the output in each pixel is corresponding to its input (Long et al., 2014). Based on FCNs, a large number of CNNs for semantic segmentation have emerged.

Among these CNNs, U-Net, proposed by Ronneberger et al. (2015), is one of the most widely used models in different areas of research. It is based on a symmetric encoder-decoder composition – in the encoder part, the size of images decreases, and the number of bands increases, while in the decoder part, the opposite direction is performed with the information generated in the encoder part being considered. In addition to the basic architecture, many variants can be applied on U-Net. For example, the encoder part of U-Net can be replaced with other models, such as Visual Geometry Group (VGG) (Simonyan & Zisserman, 2014) and Residual neural network (ResNet) (He et al., 2015). VGG was introduced to see the effect of depth on the accuracy of classification. Although a deeper network improves performance, it takes a long time to converge. Moreover, VGG has the problem of vanishing gradient issue as the model gets deeper. This problem was resolved by ResNet which introduces the so-called "identity shortcut connection" to skip one or more layers for deep networks. This technique ensures that the deeper network performs at least the same as shallow ones, not worse. The skipping also greatly reduces the training time and increases accuracy, as once a feature is learned, the model will not learn it again but rather focus on other features. There is plenty of research on satellite image segmentation or classification using these CNN models (Albert et al., 2017; Huerta et al., 2021; Pritt & Chern, 2020; Xu et al., 2020). Apart from these classical models, many other state-of-the-art models have been emerging and tested. For instance, a DeepLabv3plus model was used to extract greenspace from Gaofen-2 satellite images (Liu et al., 2019). Hartling et al. (2019) classified urban tree species by DenseNet and got higher accuracy than SVM and RF. Furthermore, High-Resolution Net (HRNet) and SegNet were tested to divide urban vegetations into evergreen trees, deciduous trees, and grassland (Xu et al., 2020).

2.3 Transfer Learning

In addition to building a CNN model from basing structure and using a large volume of training and test data, many studies relied on transfer learning from weights of already trained models on different image datasets to improve model performance and accelerate convergence due to the difficulty of obtaining labeled data (Palacios Salinas et al., 2021). These pretrained models can be roughly divided into two groups – pretrained on natural images (such as ImageNet) and remote sensing images (such as EuroSAT, BigEarthNet). Many studies relied on natural images because of their easy accessibility. However, due to the different structures of satellite images and natural images, fine-tuning models pretrained on satellite images tends to have better accuracy (Albert et al., 2017; Palacios Salinas et al., 2021). Huerta et al. (2021) used the ResNet-34 and ResNet-50 pretrained on ImageNet as the encoder part of U-Net model when performing UGS segmentation on VHR images. Albert et al. (2017) fine-tuned models pretrained on natural (ImageNet) and remote sensing images (DeepSat), and found that pretrained models based on the remote sensing benchmark yielded the best performance when doing land use classification. Due to the limited amount of labeled data and computational demand, in this research, a pretrained model was also used for the transfer learning approach.

3. Data & Methods

In section 3, the data and methods used in this research are discussed. For the data part, we include the source of datasets and data processing techniques. In the method part, the architectures and hyperparameters of models are covered.

3.1 Study Areas

Totally, we have 13 cities for model training, and 3 external cities for testing and prediction. The number of cities could have been more, but due to computational limitations, we selected relatively manageable numbers of cities. To make the model more generalized, the 13 cities for training were chosen in different continents with different latitudes. The chosen cities are San Francisco, Seattle, Denver, Philadelphia, Greater Manchester, Dublin, Amsterdam, Ghent, Dhaka, Vancouver, Dallas, London, and Buffalo – six from the USA, five from West Europe, one in Canada, and one in Asia. The

spatial extents of these cities are shown in Figure 1. Moreover, due to diverse city sizes, city structures, and population densities, the sizes of single UGS vary in these cities. For instance, the UGSs in Dhaka are mainly small ones, while in Amsterdam and some North American cities, the sizes are larger. Research has demonstrated that city structures and the layouts of objects would influence classification accuracy (Albert et al., 2017). Therefore, taking these into account can help improve model generalization.

In addition to training, we also used three external cities for validation and prediction, these are Washington D.C., Kampala, and Tel Aviv – one USA city, one African city, and one Mediterranean city. Among them, Washington D.C. and Tel Aviv were used for validation, and Kampala was used to do predictions.

3.2 Datasets

Satellite images were generated from Sentinel-2, downloaded at EarthExplorer[2], with spectral information of blue (B2), green (B3), red (B4), near-infrared (NIR, B8), and short-wave infrared (SWIR, B12). Spatial resolution for B2-B4 and B8 is 10m, and for B12 it is 20m. The central wavelengths of the chosen bands are 490nm, 560nm, 665nm, 842nm, and 2190nm, respectively. To fully delineate the green spaces in study areas, only satellite images generated between June and October 2020 with the lowest cloud cover were downloaded. Examples of downloaded satellite images are shown in Figure 1.

The ground truth datasets for UGS were obtained from the Protected Areas Database of the United States (PAD-US) 2.1 (Prior-Magee et al., 2020) for all US cities, OS Mastermap Greenspace Layer (Ordnance Survey, 2020) for UK cities. The UGS data for other cities were retrieved from open-source databases for that city respectively (City of Amsterdam, 2020; City of Ghent, 2020; City of Vancouver, 2020; Detailed Area Plan, 2020; Smart Dublin, 2020; TLV OpenData, 2020). In addition to UGS, we also added the landcover data (World Cover, 2020) to help improve model performance. The boundaries for cropping satellite images and the administrative urban boundaries for each city were obtained from Open Steet Map (OSM).

[2] https://earthexplorer.usgs.gov/

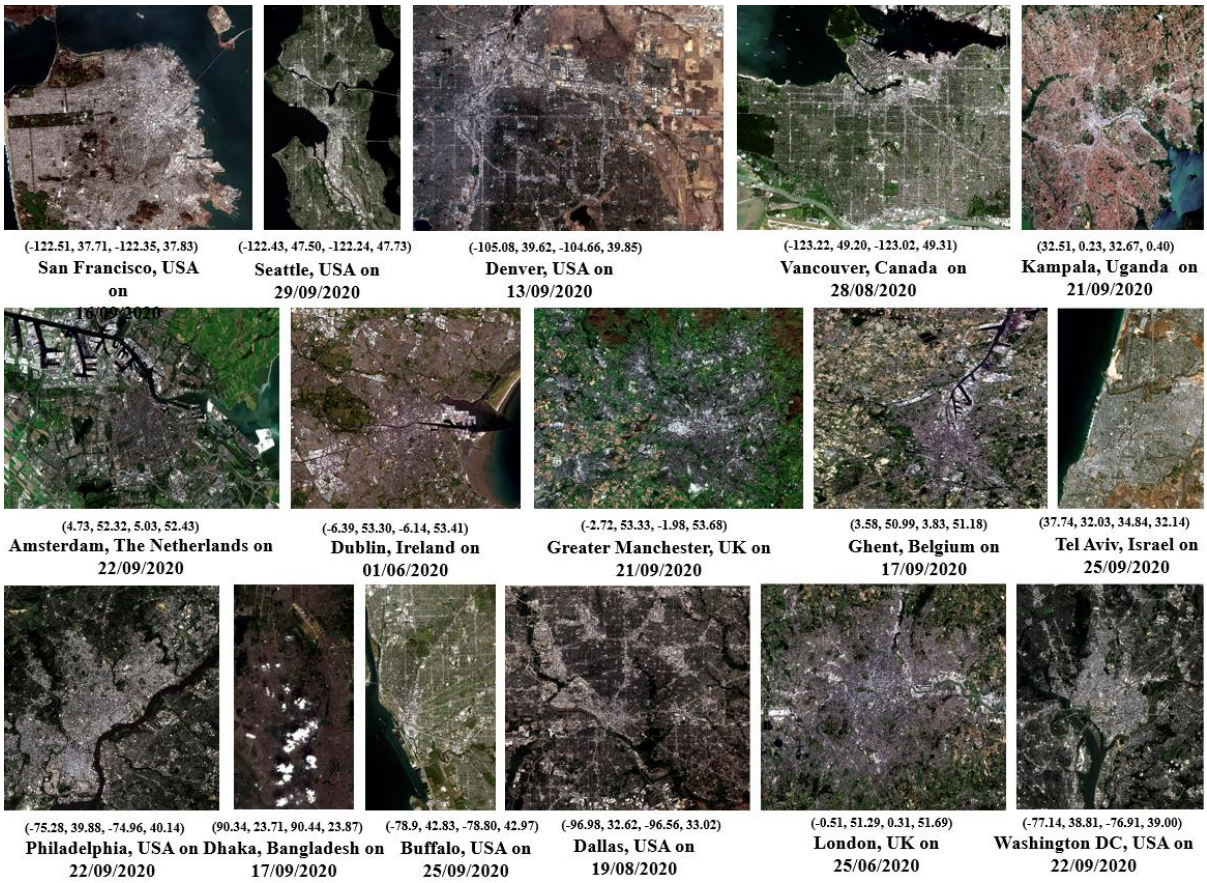

Figure 1: Examples for satellite images downloaded with generated dates (the numbers within parenthesis are the spatial extents of each city in WGS 84 coordinate system)

3.3 Data Preprocessing

To make the data structures fit the CNN models, data preprocessing is necessary. In this section, the steps of data preprocessing are discussed in detail, including indices calculation, creating image chips, and data augmentation.

3.3.1 Indices Calculation

From satellite image bands, indices can be calculated. NDVI is calculated from red (B4) and NIR (B8), with the formula shown in Equation (1). NDVI is between -1 and 1. In our research, since we only care about green spaces, and the NDVI for live plants is between 0 and 1, negative NDVIs were replaced with 0.

$$NDVI = \frac{NIR-Red}{NIR+Red} \quad (1)$$

Normalized Difference Water Index (NDWI) is calculated by

$$NDWI = \frac{Green - NIR}{Green + NIR} \quad (2)$$

where Green represents the green channel (B3). With a value between -1 and 1, NDWI is used to extract water bodies. Considering that many green spaces have lakes or rivers inside, NDWI is also widely applied in the identification of UGS (Huerta et al., 2021; Zylshal et al., 2016).

Similar to NDVI and NDWI, Normalized Difference Built-up Index (NDBI) uses NIR and SWIR to emphasize built-up areas, which is calculated from

$$NDBI = \frac{SWIR - NIR}{SWIR + NIR} \quad (3)$$

As there are only limited building structures in UGS, the NDBI of UGS should be lower than other urban areas. In this study, because the resolution of SWIR (B12) is 20m, it was resampled to 10m before being used to calculate this index.

3.3.2 Create Image Chips

Satellite images need to be cut into slices before training. 128 pixels×128 pixels, 256 pixels×256 pixels, and 512 pixels×512 pixels are the regular image sizes when conducting image segmentation (Liu et al., 2019). According to experiments of different sizes, 256 pixels×256 pixels was chosen for further analysis due to it higher accuracy than 128 pixels×128 pixels and less RAM needed than 512 pixels×512 pixels.

To generate enough image chips, stride, which represents the distance traveled during the motion, were also considered to create overlapping of chips. To let each training city has a similar contribution to the model performance, different strides were set for different cities. Stride = 32 pixels was set for San Francisco, Seattle, Denver, Ghent, Amsterdam, Dhaka, Philadelphia, Dublin, Vancouver, and Buffalo. Stride = 64 pixels was set for Greater Manchester, and stride = 80 pixels was for London and Dallas. Corresponding ground truth of UGS was first rasterized and then cut into slices using the same parameters. Examples of pairs of chips are in Figure 2.

In addition, as our research focuses on UGS, we deleted the image chip pairs if more than 50% of

their pixels are outside the administrative urban boundary. Moreover, because in urban areas only a small proportion of land is labeled as green space, the training data is imbalanced. To get balanced data to improve accuracy (Lu et al., 2006), we removed the image chips whose proportion of backgrounds (non-UGS pixels) is more than a certain threshold (Liu et al., 2019). According to experiments on the trade-off between the training set size and degree of balance, the thresholds were set to be 90% for Dhaka and 86% for other cities[3]. The total image chip pairs we got for training were 4299, with the number of pixels of green space 71.9 million, and backgrounds 209.9 million. All layers of images are normalized to the same range of zero and one before training.

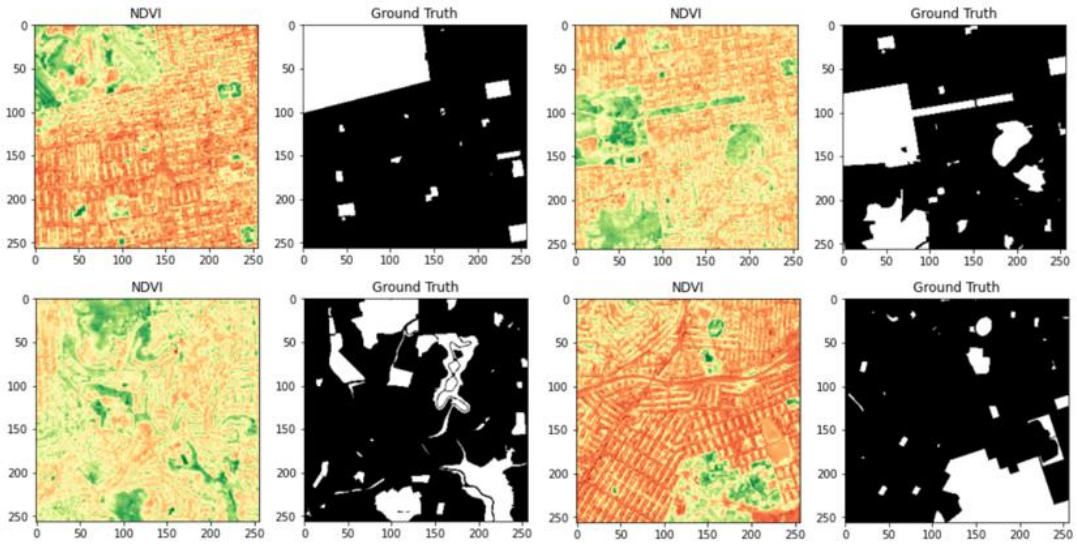

Figure 2: Examples of image chip pairs (to make the green spaces explicit, NDVI was plotted, with green pixels representing green areas. For ground truth images, the white areas are UGSs)

3.3.3 Data Augmentation

Data augmentation was done to increase the size of the training set. Techniques chosen were (1) random rotation within an angle of 45°; (2) random width and height shift within a range of 20%; (3) randomly horizontal and vertical flip; (4) randomly zooming in and out within a range of 20%. These techniques were applied to satellite image chips and corresponding masks, with the generated images being put into training to improve model performance (Liu et al., 2019). Furthermore, the generated images can be seen as noise added to the original training samples, helping reduce the overfitting of

[3] Threshold for Dhaka was set higher because green spaces in Dhaka are mainly small ones. If setting the threshold to be 86%, few chips were left.

the CNN models (Shorten & Khoshgoftaar, 2019). Examples of images produced by data augmentation are plotted in Figure 3.

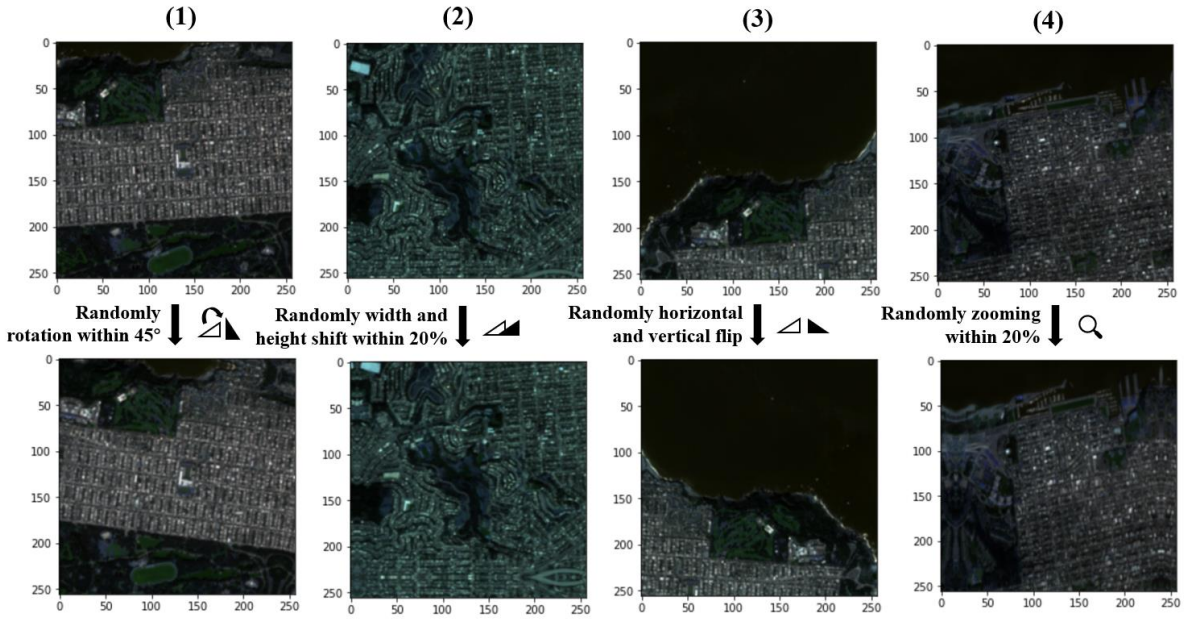

Figure 3: Examples of generated images by data augmentation algorithms

3.4 Model Setup

The U-Net model was chosen to do the segmentation for UGS. Both U-Net models from base level and using pretrained models as backbones have been tested. All model setup was done using python packages of Keras and TensorFlow in Google Colab with a RAM of 25.46 GB.

3.4.1 U-Net Model

The classical encoder-decoder architecture of U-Net model was implemented in our study. The encoder part consists of a sequence of convolutions, each followed by an activation function, which is set as rectified linear unit (ReLU) according to other studies (Chen et al., 2021; Huerta et al., 2021), and a max-pooling operation. In this part, spatial information of images is reduced while feature information increases. After reaching the bridge, the decoder part combines the spatial and feature information through sets of up-convolutions and concatenation with information generated from the corresponding block of the encoder part (Ronneberger et al., 2015).

In our case, due to the limited computational capacity, different 3-channel combinations from the total

8 layers were tested. Based on the similar research of Huerta et al. (2021), the chosen combinations were (1) Red-Green-Blue; (2) Red-Green-NIR; (3) NDVI-Red-NIR; (4) NDWI-Red-NIR; (5) NDBI-Red-NIR; (6) NDVI-NDWI-NDBI; (7) NDVI-NDWI-landcover; (8) NDVI-NDBI-landcover.

The model architecture used in our study is presented in Figure 4. The size of the input image is 256×256×3, after 4 encoder blocks, the feature size becomes 32×32×512. Then, it comes to the bridge. Different from the encoder blocks, no max-pooling is conducted here. The opposite operations are performed in the decoder part until we get the size of 256×256×64. The final operation is a 1×1 convolution, with the feature size of 256×256×1 as the output. The reason for choosing 1×1 convolution is that we are doing a binary segmentation (i.e., UGS or not).

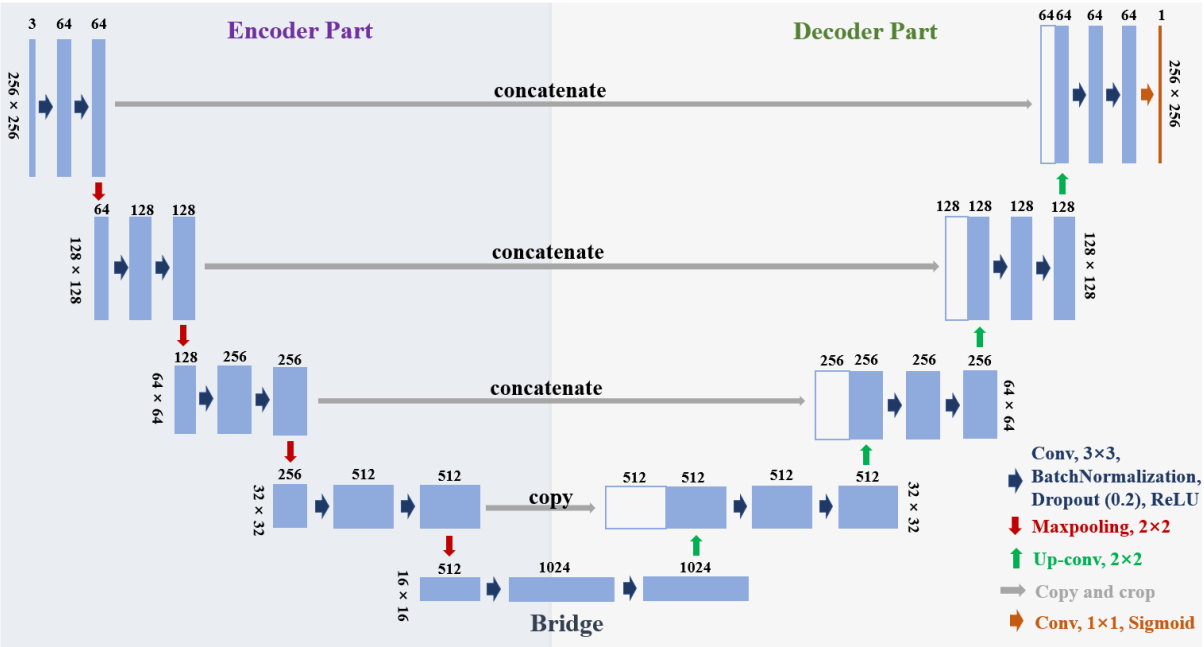

Figure 4: Architecture of the U-Net model

3.4.2 U-Net Model with Pretrained Backbones

In addition to U-Net model from scratch or base level, pretrained backbones were also used to improve accuracy and accelerate training. ResNet-50 pretrained on ImageNet (He et al., 2015) was chosen to replace the encoder part of the U-Net model in 3.4.1, with its architecture explained in Figure 5. The ResNet-50 in the encoder part consists of 4 stages, each stage includes 1 Conv block and 2, 3, 5, and 2 Identity blocks respectively. The Conv block and Identity block are the repetitions of Convolution, BatchNormalization, and ReLU operations, with details shown at the bottom right of

Figure 5. The decoder part of this model is the same as in 3.4.1. The only notable thing is that the size of the output image would be 256×256×2 instead of 256×256×1. The reason for this is that we are using one-hot encoding here, so the ground truth chips were first categorized before being put into training.

To contrast with ResNet-50, a pretrained VGG-16 from ImageNet (Figure 6) was also adopted as the encoder part of the U-Net model. Instead of the two convolutional operations in each encoder block, VGG-16 has 2 or 3 convolutional operations in each block.

To make the models with backbones comparable with the scratch model, the same 3-band combinations as in 3.4.1 were applied.

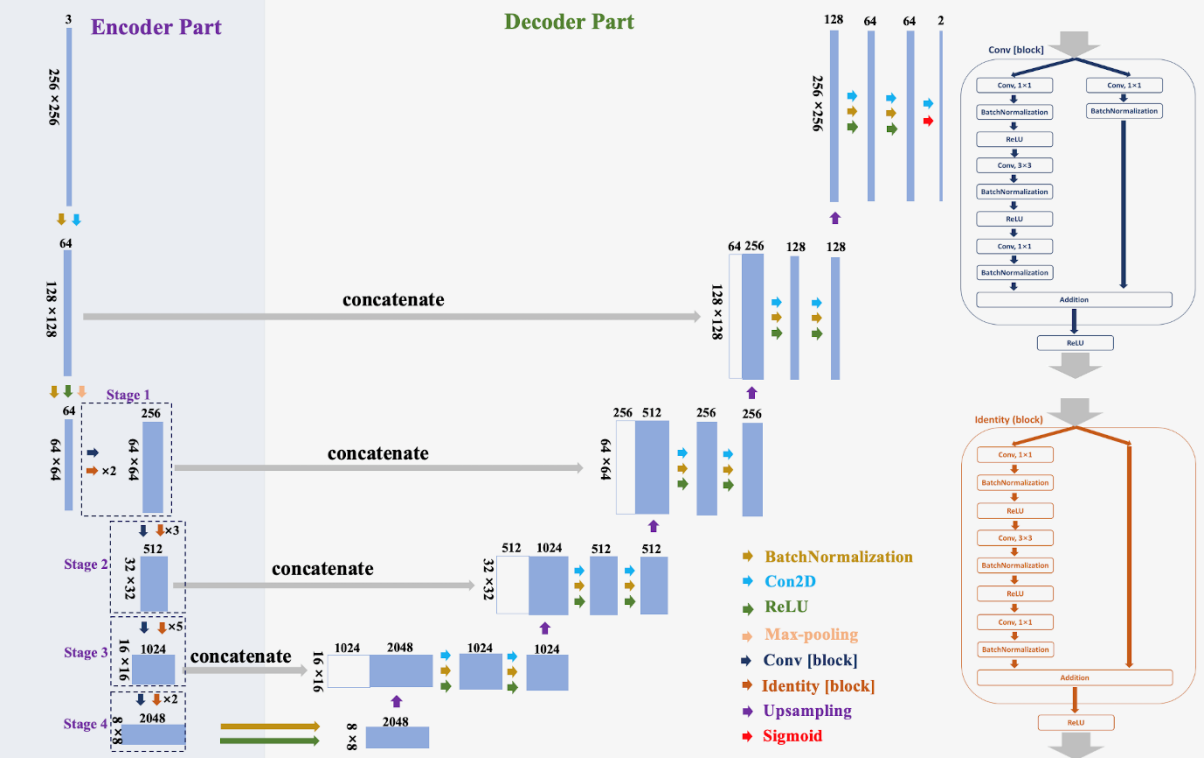

Figure 5: Architecture of U-Net model with ResNet-50 backbone

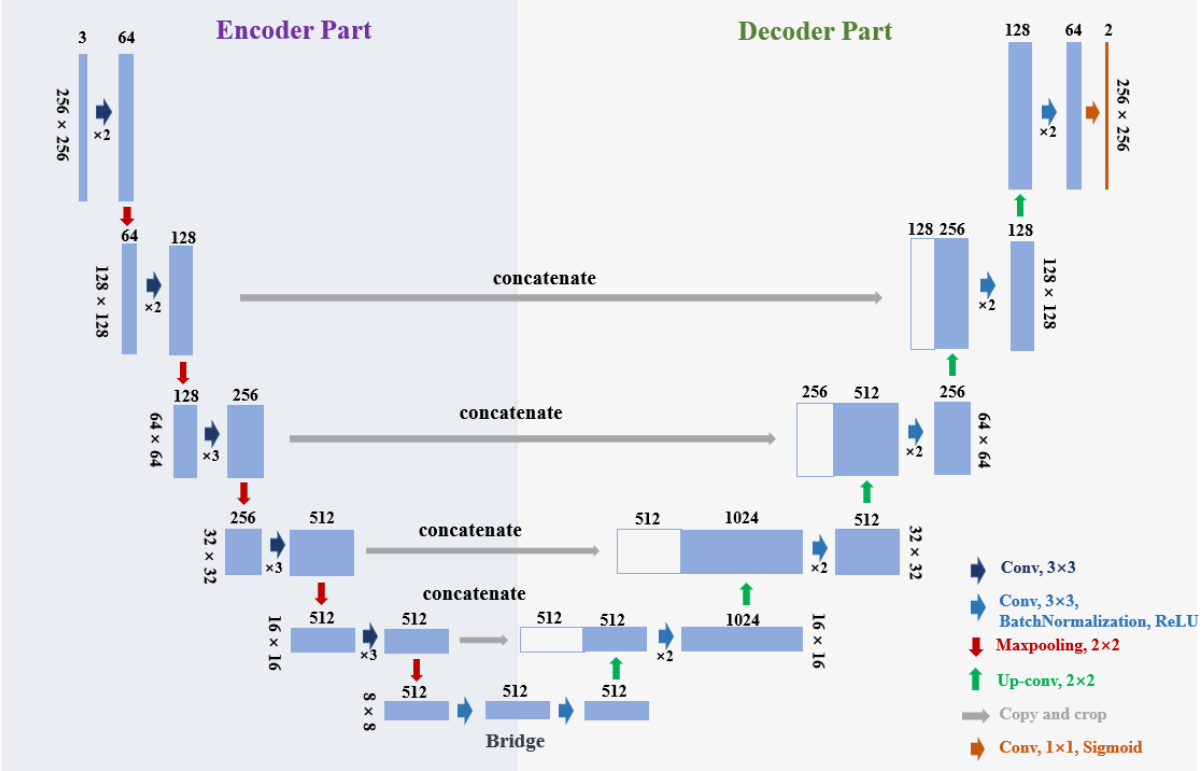

Figure 6: Architecture of U-Net model with VGG-16 backbone

3.4.3 Metrics, Losses & Other Hyperparameters

Due to the issue of imbalanced classes, apart from the overall accuracy (OA) – the sum of the true positives (TPs) plus true negatives (TNs) divided by the total number of cells tested (Alberg et al., 2004), Intersection over Union (IoU) (Rahman & Wang, 2016; Xu et al., 2020) and F-score (Huerta et al., 2021) were also chosen as the metrics.

IoU, also known as Jaccard Index, is calculated as the size of intersection divided by the size of union:

$$J(A,B) = \frac{A \cap B}{A \cup B} \tag{4}$$

The higher the overlap, the better the segmentation. F-score (Dice Coefficient) is the harmonic average of precision and recall and is calculated by

$$F_\beta(precision, recall) = (1+\beta^2)\frac{precsion \cdot recall}{\beta^2 \cdot precision + recall} \tag{5}$$

where β is the coefficient to balance precision and recall, and in this research, β is set to be one. Similar to IoU, F-score is best at one and worst at zero. AUC (Area Under the ROC curve) was also

reported to provide an aggregate measure of performance across all possible classifications. AUC ranges between zero and one, with the value of a model having 100% correct classifications to be one.

The loss function used was Binary Focal Loss. It penalizes the major class and thus is more suitable for an imbalanced training set. The formula to calculate is

$$L(gt, pr) = -gt \cdot \alpha \cdot (1 - pr)^{\gamma} \cdot \log(pr) - (1 - gt) \cdot \alpha \cdot pr^{\gamma} \cdot \log(1 - pr) \quad (6)$$

where gt represents ground truth, pr is the prediction, α is the weighting factor in balanced classification, γ is the modulating factor used to penalize the majority. In this study, α was chosen to be 0.25 and γ to be 2.

85% of the dataset was randomly sampled to be the training set, and 15% to be the validation set. To make the training more stable, the learning rate was chosen as 0.0001, with a batch size of 16. The total epochs for training were set to be 50. Due to the data augmentation techniques used, the training steps per epoch were calculated as the size of the training set divided by the batch size, which is 228, and the validation steps per epoch were the size of the validation set divided by the batch size, which is 40.

4. Results

Model performance and predictions on external cities are discussed in this section. The results of both U-Net model from scratch and with pretrained backbones are included.

4.1 Semantic Segmentation of UGS in Training Cities

Performance of U-Net model from scratch or base level is documented in Table 1. The average OA achieved from different combinations of bands was 0.8624, with IoU 0.6259, F-score 0.6645, and AUC 0.7640. Among all the combinations, NDWI-Red-NIR and NDVI-NDBI-landcover gave the best performances, while the accuracy got from Red-Green-Blue was low. From the results, we can generally conclude that adding indices into original bands of remote sensing images is beneficial to obtain better performance, whereas the choosing and combination of indices need to be done carefully. For instance, in our case, the combination of NDVI and NDWI made the model not function well,

however, NDWI together with Red and NIR achieved the highest accuracy. This is understandable since the combination of Red and NIR gives the information of NDVI. Along with NDWI, this can provide a depiction of UGSs with water bodies inside. However, adding another index, such as NDBI or landcover, introduced redundancy due to the correlation among these variables.

Table 1: Performance of U-Net model from base level

Combination of bands	OA	IoU	F-score	AUC
Red-Green-Blue	0.8329	0.5394	0.5438	0.6894
Red-Green-NIR	0.8531	0.6425	0.6227	0.7373
NDVI-Red-NIR	0.8724	0.6429	0.6921	0.7745
NDWI-Red-NIR	0.8977	0.7185	0.7819	0.8457
NDBI-Red-NIR	0.8498	0.6072	0.6583	0.7599
NDVI-NDWI-NDBI	0.8515	0.5846	0.6126	0.7252
NDVI-NDWI-landcover	0.8502	0.5964	0.6365	0.7431
NDVI-NDBI-landcover	0.8914	0.7044	0.7682	0.8368
Average	0.8624	0.6295	0.6645	0.7640

The performance of U-Net model with backbones is shown in Table 2. ResNet-50 backbone added benefits to the model, while VGG-16 made the model perform worse. When using the ResNet-50 backbone, the average OA was 0.9695, the average IoU was 0.9188, the average F-score was 0.9570, and the average AUC was 0.9584, while under the VGG-16 backbone, the metrics were 0.5139, 0.2515, 0.3303, 0.4972 respectively. This is similar to Albert et al. (2017), where the authors achieved higher accuracy when adopting ResNet than VGG architectures.

Different combinations of bands gave very similar performance when using ResNet-50 backbone, indicating its good learning skill across different combinations of inputs. However, it showed a large difference when using the pretrained VGG-16 backbone. Higher performance was observed from the combinations of Red-Green-NIR, NDVI-Red-NIR, NDBI-Red-NIR, and NDVI-NDWI-landcover, however, only about 20% of OA and F-score were achieved using Red-Green-Blue, NDWI-red-NIR, NDVI-NDWI-NDBI, and NDVI-NDBI-landcover. This indicated the limited ability of the VGG-16 backbone on segmentation of UGSs.

Table 2: Model performance of U-Net model with backbones

Combination of bands	ResNet-50				VGG-16			
	OA	IoU	F-score	AUC	OA	IoU	F-score	AUC
Red-Green-Blue	0.9708	0.9220	0.9590	0.9610	0.2535	0.1243	0.2001	0.5013
Red-Green-NIR	0.9718	0.9243	0.9602	0.9596	0.7508	0.3769	0.4319	0.4996
NDVI-Red-NIR	0.9709	0.9223	0.9591	0.9608	0.7259	0.3857	0.4588	0.4952
NDWI-Red-NIR	0.9712	0.9225	0.9592	0.9616	0.2804	0.1456	0.2409	0.5063
NDBI-Red-NIR	0.9704	0.9208	0.9583	0.9607	0.7439	0.3766	0.4348	0.4980
NDVI-NDWI-NDBI	0.9691	0.9178	0.9566	0.9565	0.2476	0.1222	0.1960	0.5000
NDVI-NDWI-landcover	0.9660	0.9107	0.9526	0.9532	0.6020	0.3581	0.4834	0.4768
NDVI-NDBI-landcover	0.9656	0.9100	0.9522	0.9536	0.2468	0.1222	0.1961	0.5000
Average	0.9695	0.9188	0.9572	0.9584	0.5139	0.2515	0.3303	0.4972

We also investigated the training histories of different models. Because the same model generated similar histories, the converge history of only one combination (NDVI-NDWI-NDBI) is drawn in Figure 7. The best convergence was achieved by the ResNet-50 backbone. After training for about 6 epochs, the loss of the model with the ResNet-50 backbone began to converge to the minimum, while the VGG-16 backbone had difficulty converging within the 50 epochs. Additionally, judging from the training and validation loss, overfitting can be detected from scratch model and model with VGG-16 backbone. This again demonstrates their limitation in UGS segmentation. Thus, U-Net with pretrained ResNet-50 was chosen to do further analysis.

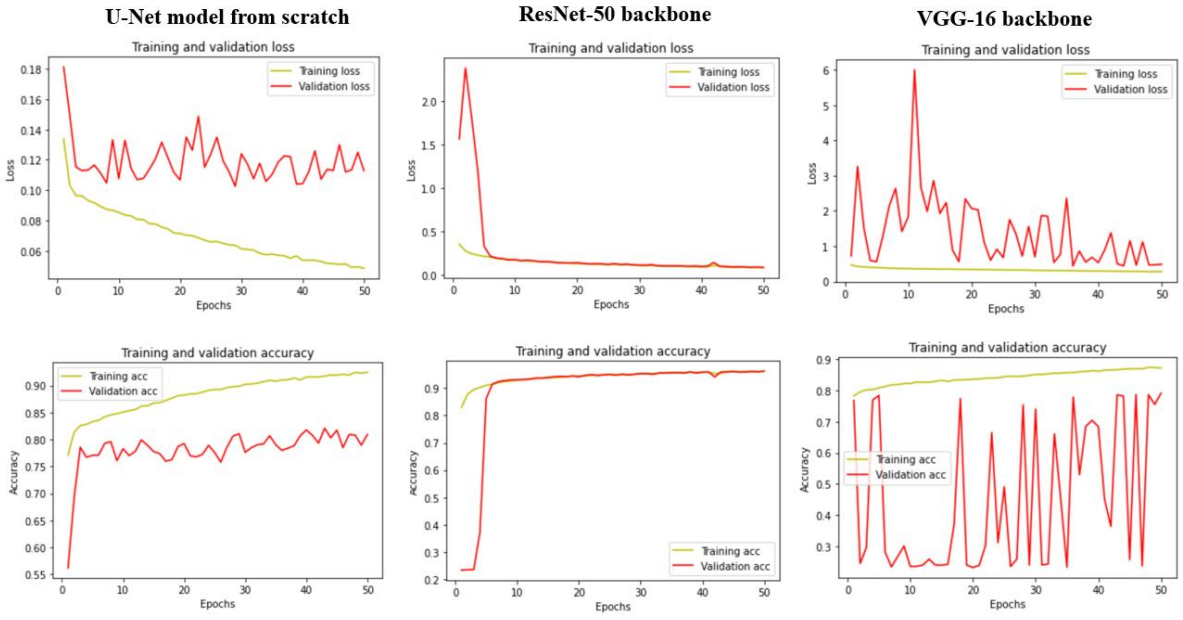

Figure 7: Training history (upper: loss, lower: accuracy) of U-Net models using NDVI-NDWI-NDBI

4.2 Validation and Prediction on External Cities

Validations on Washington D.C. and Tel Aviv of U-Net model with the ResNet-50 backbone are shown in Table 3. The average OA, IoU, F-score, and AUC for Washington D.C. are 0.8743, 0.6185, 0.7236, and 0.7429, and for Tel Aviv, the metrics are 0.8790, 0.4954, 0.5660, and 0.5639. This showed a moderate to good generalization capacity of our model. In general, the model achieved better performance on Washington D.C. than on Tel Aviv. This is reasonable since a high proportion of the training set were US and European cities. They have similar structures to Washington D.C., while Tel Aviv shows a different UGS appearance compared with the training cities (Albert et al., 2017).

Different combinations have various strengths when detecting UGSs, therefore we averaged all the combinations to get the final predicted maps (Figure 8). Intersections of prediction and ground truth are plotted in green, predicted UGSs while not in the ground truth set are shown in purple, and ground truth UGSs which are not identified are in blue. As shown in the plots, Washington D.C. has better accuracy than Tel Aviv, consistent with the metrics above. In the western part of Washington D.C., the model has better performance, whereas in the eastern part, our model tends to extract more areas as UGSs. In Tel Aviv, most green spaces are small, and our model has limited accuracy to identify these small areas. The prediction for Kampala was also plotted.

To get a better picture of model performance, we also used Washington D.C. as an example to explore where our model performed well and where could be improved, with results shown in Figure 9. Our model made a good prediction for clumped UGSs, such as the Rock Creek Park and East Potomac Park circled in red. However, for UGSs which have buildings inside, for instance, the blue areas in rectangle, our model did not give an accurate identification. Additionally, the model also generated some UGSs which are not in the ground truth dataset. These areas are mainly golf courses, national parks, and cemeteries, which are not freely accessible to the public or not normal UGSs under the hard definition of UGS (Nordh et al., 2017).

Table 3: Performance of U-Net model with ResNet-50 backbone of test cities

Combination of bands	Washington D.C.				Tel Aviv			
	OA	IoU	F-score	AUC	OA	IoU	F-score	AUC
Red-Green-Blue	0.8714	0.5905	0.6848	0.7000	0.8540	0.4644	0.5290	0.5257
Red-Green-NIR	0.8612	0.5905	0.7034	0.7070	0.8901	0.4933	0.5549	0.5383
NDVI-Red-NIR	0.8776	0.6529	0.7601	0.8061	0.9110	0.5112	0.5766	0.5653
NDWI-Red-NIR	0.8773	0.5769	0.6779	0.6708	0.8843	0.5182	0.6003	0.5965
NDBI-Red-NIR	0.8752	0.6257	0.7309	0.7509	0.8749	0.4865	0.5534	0.5436
NDVI-NDWI-NDBI	0.8837	0.6517	0.7584	0.7774	0.8104	0.4500	0.5292	0.5600
NDVI-NDWI-landcover	0.8861	0.6368	0.7411	0.7578	0.8963	0.5337	0.6177	0.6271
NDVI-NDBI-landcover	0.8615	0.6233	0.7325	0.7732	0.9108	0.5059	0.5671	0.5545
Average	0.8743	0.6185	0.7236	0.7429	0.8790	0.4954	0.5660	0.5639

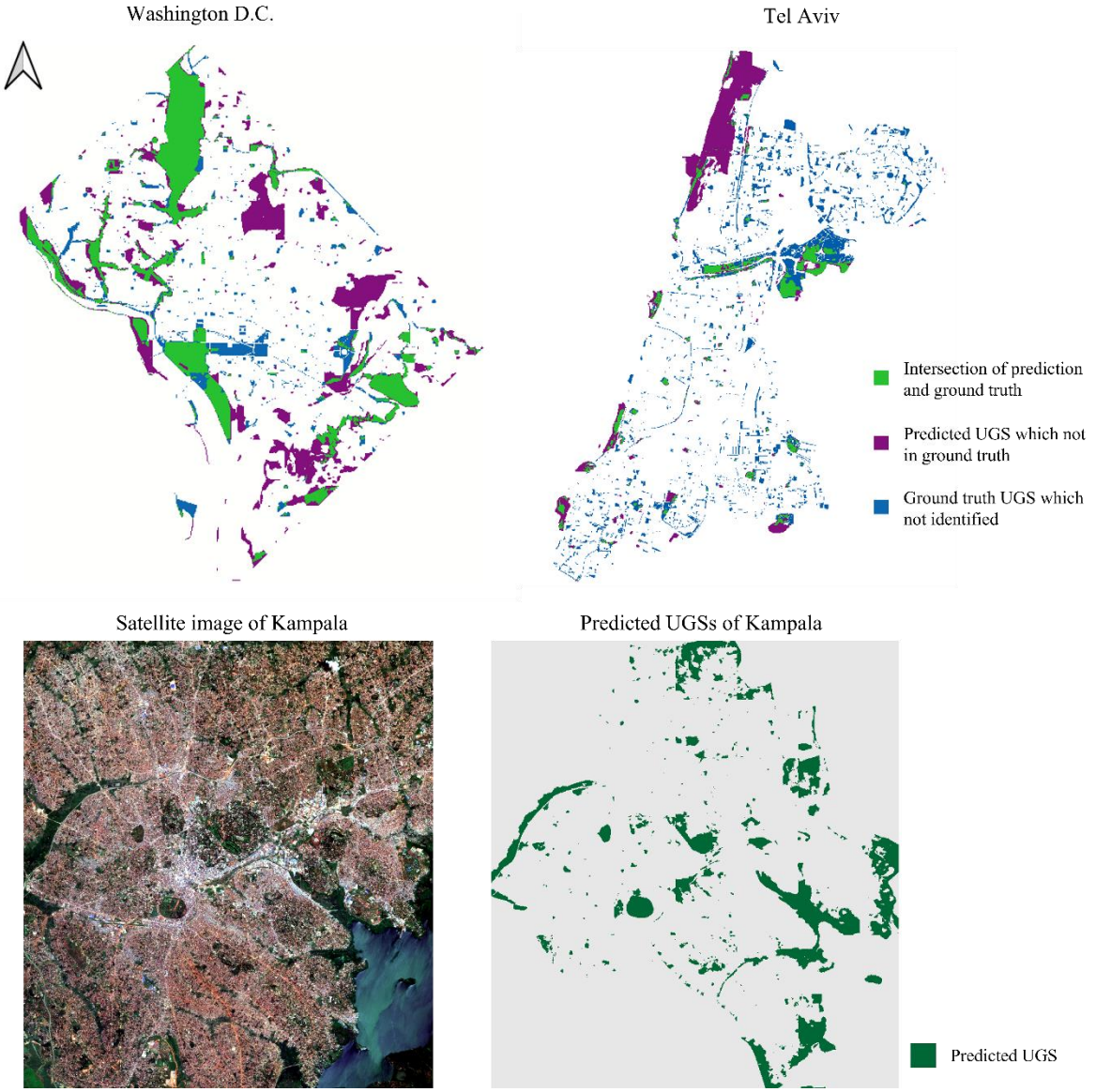

Figure 8: upper: Ground truth and predicted UGSs for Washington D.C. and Tel Aviv, lower: satellite image and predicted UGSs for Kampala

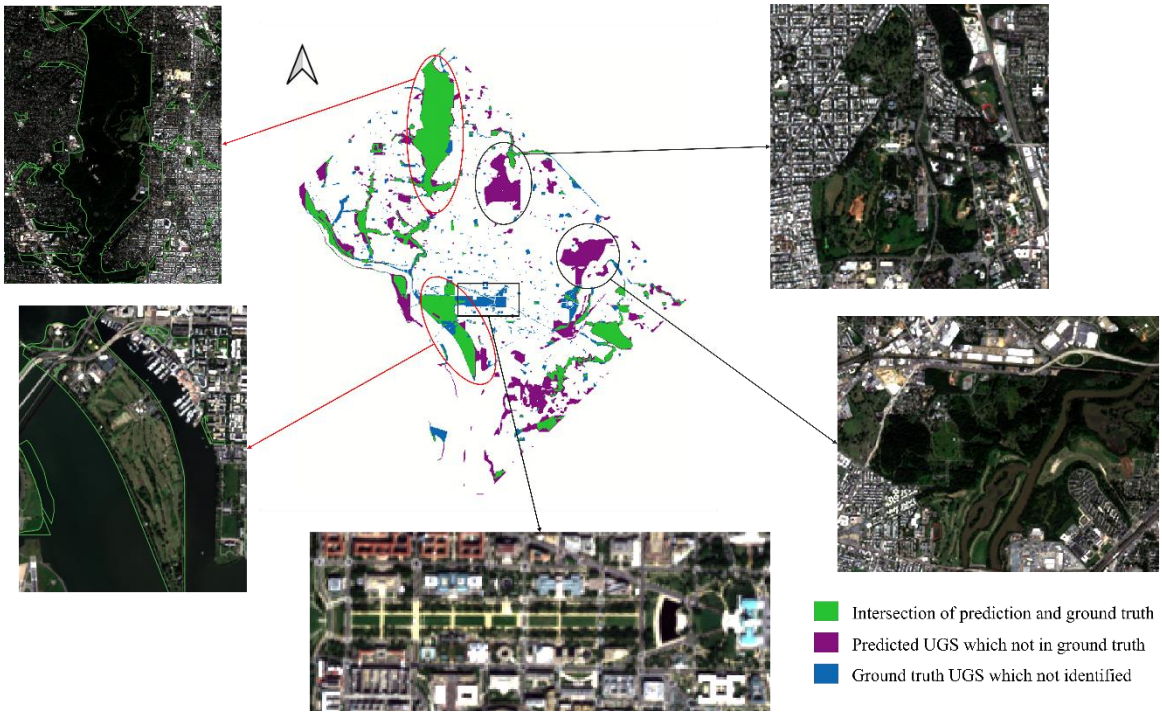

Figure 9: Model assessment of Washington D.C. (areas in circles or rectangles are augmented and shown as the satellite images)

5. Discussion

This paper implemented semantic segmentation for public UGSs using satellite images and CNN models. Compared with similar research, our study has a much more diverse choice of study areas, helping the model to have a sufficient performance when using moderate resolution data. Meanwhile, most related research did not use different study areas for training and testing while only extracting some areas within the same city for validation, such as Albert et al. (2017), and Huerta et al. (2021), and Liu et al. (2019). This paper, however, explored the model performance on external cities to check the generalization ability of models, thus it is more practical if aiming to use the models to identify UGSs for cities where benchmark city data is not available, for instance, cities in the global south. Additionally, these models can also be useful to update the UGS data by re-running the models on new images in the future.

Moreover, our model generated many UGSs which were not included in the ground truth datasets, such as golf courses, national parks, and cemeteries. However, some scholars suggested that these places should be covered in UGS datasets as they provide the same or even more benefits (e.g.,

reflection and memory functions of cemeteries) UGSs can provide for the citizens (M'Ikiugu et al., 2012; Skår et al., 2018). Therefore, this paper might have given a potential method to get a more complete dataset for UGS distributions.

The satellite images used in this research were moderate resolution ones. We chose these because they are freely accessible, despite it may add a limitation to the model performance in comparison with other studies that used VHR images (Hartling et al., 2019; Huerta et al., 2021; Liu et al., 2019; Xu et al., 2020). If using images with higher resolution, the model performance could be increased further.

Limitation of hardware has often been a challenge when training CNNs, and our research may give some insights into this. Making use of pretrained weights is a good option when you could only put a limited size of labeled data into models (Palacios Salinas et al., 2021). In this study, two backbones – ResNet-50 and VGG-16 pretrained on ImageNet were used. Compared with the model from scratch, the ResNet-50 backbone improved the model performance a lot and made the training converge faster. Additionally, to deal with the huge data size caused by multiple satellite image layers, different combinations of bands were used instead.

This research undoubtedly has limitations. First, due to data availability and limited system memory, only 13 cities were pooled for training, with a high proportion of US and European cities. Therefore, the training set might not have a comprehensive delineation of UGSs in different regions, limiting the generalization capacities of the model on external cities. Second, the backbones chosen for this study were pretrained on natural images. As mentioned before, remote sensing images have a different structure than natural images, so using weights pretrained on satellite images might yield better results. Third, only U-Net model was implemented, however, considering more architectures might bring other possibilities.

On basis of this, we proposed the potential improvements for future research as follows: first, more cities (hundreds or even thousands) can be put into the CNN models with the help of high-performance computing to boost model generalization. Second, using the segmented images as a fundament, further object-based image analysis (OBIA) (Zylshal et al., 2016) or contextual guided models (Zhao et al., 2017) can be applied to improve accuracy and to make the boundaries of detected objects more similar to visual interpretations.

6. Conclusion

In this paper, we aimed to build a CNN model to automatically detect UGSs from Sentinel-2 satellite images. To achieve this, we used 13 cities on different continents for training and 3 external cities for testing. The remote sensing images downloaded were first cropped into proper size (256 pixels×256 pixels) before being put into models. Data augmentation was also done to affluent and add variants to the original dataset, helping increase accuracy and reduce overfitting. Eight three-band combinations chosen from the layers of satellite images and calculated indices were put into both U-Net model from base level and U-Net model with pretrained backbones (ResNet-50 and VGG-16 pretrained on ImageNet). U-Net model with pretrained ResNet-50 achieved the highest performance and thus was used to do validation and prediction for external cities. Results showed our model to be a good digital tool for extracting UGSs from satellite images – the average IoU and F-score achieved for Washington D.C. were 0.6185 and 0.7236, and for Tel Aviv, the metrics were 0.4954 and 0.5600 respectively.

On top of our study, improvements and further research can be implemented. Pooling more cities into training might give a more comprehensive delineation of UGSs in different regions and boost model accuracy. Then, this model can be used to predict UGSs for global some southern cities, where the benchmark datasets are not available. Furthermore, other studies based on UGS distribution, such as UGS accessibility, and causal research between UGS and socio-demographical variables can be conducted.

Acknowledgment

I would like to extend my sincere thanks to Dr. Labib S.M., the supervisor of this thesis project, for his professional and patient guidance throughout my endeavor. I also could not have undertaken this journey without my group member – Marta Kozłowska, who generously provided help when processing data and building models.

Lastly, I would be remiss in not mentioning my family and friends. Their belief in me has kept my spirits and motivation high during this process. Additionally, I will give my thanks to my cat, Jianguo Zhu, for all the entertainment and emotional support.

References

Alberg, A. J., Park, J. W., Hager, B. W., Brock, M. V., & Diener-West, M. (2004). The use of "overall accuracy" to evaluate the validity of screening or diagnostic tests. *Journal of General Internal Medicine*, *19*(5), 460–465. https://doi.org/10.1111/j.1525-1497.2004.30091.x

Albert, A., Kaur, J., & Gonzalez, M. (2017). *Using convolutional networks and satellite imagery to identify patterns in urban environments at a large scale*. https://doi.org/10.48550/ARXIV.1704.02965

Aronson, M. F., Lepczyk, C. A., Evans, K. L., Goddard, M. A., Lerman, S. B., MacIvor, J. S., Nilon, C. H., & Vargo, T. (2017). Biodiversity in the city: Key challenges for urban green space management. *Frontiers in Ecology and the Environment*, *15*(4), 189–196. https://doi.org/10.1002/fee.1480

Chen, Y., Weng, Q., Tang, L., Liu, Q., Zhang, X., & Bilal, M. (2021). Automatic mapping of urban green spaces using a geospatial neural network. *GIScience & Remote Sensing*, *58*(4), 624–642. https://doi.org/10.1080/15481603.2021.1933367

City of Amsterdam. (2020). *Maps Amsterdam*. https://maps.amsterdam.nl/?LANG=en

City of Ghent. (2020). *Explore—Open Data Portaal—Stad Gent*. https://data.stad.gent/explore/?disjunctive.keyword&disjunctive.theme&sort=modified

City of Vancouver. (2020). *Home—City of Vancouver Open Data Portal*. https://opendata.vancouver.ca/pages/home/

Detailed Area Plan. (2020). *Detailed Area Plan (DAP)*. http://43.243.207.51/Website/R2_2_1/viewer.htm

Earth Observatory. (2022, June). *Measuring Vegetation (NDVI & EVI)*. https://earthobservatory.nasa.gov/features/MeasuringVegetation

European Environment Agency. (2022, February). *Copernicus Land Monitoring Service—Urban Atlas—European Environment Agency*. https://www.eea.europa.eu/data-and-maps/data/copernicus-land-monitoring-service-urban-atlas

European Space Agency. (2022a, June). *GeoEye-1—Earth Online*. https://earth.esa.int/eogateway/missions/geoeye-1

European Space Agency. (2022, June). *QuickBird-2—Earth Online*. https://earth.esa.int/eogateway/missions/quickbird-2

European Space Agency. (2022b, June). *Spatial—Resolutions—Sentinel-2 MSI - User Guides—Sentinel Online—Sentinel Online*.

https://sentinels.copernicus.eu/web/sentinel/user-guides/sentinel-2-msi/resolutions/spatial

European Space Agency. (2022c, June). *WorldView-2—Earth Online*. https://earth.esa.int/eogateway/missions/worldview-2

Gupta, K., Kumar, P., Pathan, S. K., & Sharma, K. P. (2012). Urban Neighborhood Green Index – A measure of green spaces in urban areas. *Landscape and Urban Planning*, *105*(3), 325–335. https://doi.org/10.1016/j.landurbplan.2012.01.003

Hartig, T., Mitchell, R., de Vries, S., & Frumkin, H. (2014). Nature and Health. *Annual Review of Public Health*, *35*(1), 207–228. https://doi.org/10.1146/annurev-publhealth-032013-182443

Hartling, S., Sagan, V., Sidike, P., Maimaitijiang, M., & Carron, J. (2019). Urban Tree Species Classification Using a WorldView-2/3 and LiDAR Data Fusion Approach and Deep Learning. *Sensors*, *19*(6), 1284. https://doi.org/10.3390/s19061284

He, K., Zhang, X., Ren, S., & Sun, J. (2015). *Deep Residual Learning for Image Recognition*. https://doi.org/10.48550/ARXIV.1512.03385

Huerta, R. E., Yépez, F. D., Lozano-García, D. F., Guerra Cobián, V. H., Ferriño Fierro, A. L., de León Gómez, H., Cavazos González, R. A., & Vargas-Martínez, A. (2021). Mapping Urban Green Spaces at the Metropolitan Level Using Very High Resolution Satellite Imagery and Deep Learning Techniques for Semantic Segmentation. *Remote Sensing*, *13*(11), 2031. https://doi.org/10.3390/rs13112031

Kranjčić, N., Medak, D., Župan, R., & Rezo, M. (2019). Machine Learning Methods for Classification of the Green Infrastructure in City Areas. *ISPRS International Journal of Geo-Information*, *8*(10), 463. https://doi.org/10.3390/ijgi8100463

Labib, S. M., & Harris, A. (2018). The potentials of Sentinel-2 and LandSat-8 data in green infrastructure extraction, using object based image analysis (OBIA) method. *European Journal of Remote Sensing*, *51*(1), 231–240. https://doi.org/10.1080/22797254.2017.1419441

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444. https://doi.org/10.1038/nature14539

Liu, W., Yue, A., Shi, W., Ji, J., & Deng, R. (2019). An Automatic Extraction Architecture of Urban Green Space Based on DeepLabv3plus Semantic Segmentation Model. *2019 IEEE 4th International Conference on Image, Vision and Computing (ICIVC)*, 311–315. https://doi.org/10.1109/ICIVC47709.2019.8981007

Long, J., Shelhamer, E., & Darrell, T. (2014). *Fully Convolutional Networks for Semantic Segmentation*. https://doi.org/10.48550/ARXIV.1411.4038

Lu, X., Zheng, B., Velivelli, A., & Zhai, C. (2006). Enhancing Text Categorization with Semantic-enriched Representation and Training Data Augmentation. *Journal of the American Medical Informatics Association*, *13*(5), 526–535. https://doi.org/10.1197/jamia.M2051

Ludwig, C., Hecht, R., Lautenbach, S., Schorcht, M., & Zipf, A. (2021). Mapping Public Urban Green Spaces Based on OpenStreetMap and Sentinel-2 Imagery Using Belief Functions. *ISPRS International Journal of Geo-Information*, *10*(4), 251. https://doi.org/10.3390/ijgi10040251

Markevych, I., Schoierer, J., Hartig, T., Chudnovsky, A., Hystad, P., Dzhambov, A. M., de Vries, S., Triguero-Mas, M., Brauer, M., Nieuwenhuijsen, M. J., Lupp, G., Richardson, E. A., Astell-Burt, T., Dimitrova, D., Feng, X., Sadeh, M., Standl, M., Heinrich, J., & Fuertes, E. (2017). Exploring pathways linking greenspace to health: Theoretical and methodological guidance. *Environmental Research*, *158*, 301–317. https://doi.org/10.1016/j.envres.2017.06.028

M'Ikiugu, M. M., Kinoshita, I., & Tashiro, Y. (2012). Urban Green Space Analysis and Identification of its Potential Expansion Areas. *Procedia - Social and Behavioral Sciences*, *35*, 449–458. https://doi.org/10.1016/j.sbspro.2012.02.110

More, N., Nikam, V. B., & Banerjee, B. (2020). Machine learning on high performance computing for urban greenspace change detection: Satellite image data fusion approach. *International Journal of Image and Data Fusion*, *11*(3), 218–232. https://doi.org/10.1080/19479832.2020.1749142

Nordh, H., Evensen, K. H., & Skår, M. (2017). A peaceful place in the city—A qualitative study of restorative components of the cemetery. *Landscape and Urban Planning*, *167*, 108–117. https://doi.org/10.1016/j.landurbplan.2017.06.004

Ordnance Survey. (2020, June). *OS Open Greenspace | OS Products*. https://www.ordnancesurvey.co.uk/business-government/products/open-map-greenspace

Palacios Salinas, N. R., Baratchi, M., van Rijn, J. N., & Vollrath, A. (2021). Automated Machine Learning for Satellite Data: Integrating Remote Sensing Pre-trained Models into AutoML Systems. In Y. Dong, N. Kourtellis, B. Hammer, & J. A. Lozano (Eds.), *Machine Learning and Knowledge Discovery in Databases. Applied Data Science Track* (Vol. 12979, pp. 447–462). Springer International Publishing. https://doi.org/10.1007/978-3-030-86517-7_28

Prior-Magee, J. S., Johnson, L. J., Croft, M. J., Case, M. L., Belyea, C. M., & Voge, M. L. (2020).

Protected Areas Database of the United States (PAD-US) 2.1 (Provisional Release [Data set]. U.S. Geological Survey. https://doi.org/10.5066/P92QM3NT

Pritt, M., & Chern, G. (2020). *Satellite Image Classification with Deep Learning*. https://doi.org/10.48550/ARXIV.2010.06497

Rahman, M. A., & Wang, Y. (2016). Optimizing Intersection-Over-Union in Deep Neural Networks for Image Segmentation. In G. Bebis, R. Boyle, B. Parvin, D. Koracin, F. Porikli, S. Skaff, A. Entezari, J. Min, D. Iwai, A. Sadagic, C. Scheidegger, & T. Isenberg (Eds.), *Advances in Visual Computing* (Vol. 10072, pp. 234–244). Springer International Publishing. https://doi.org/10.1007/978-3-319-50835-1_22

Ronneberger, O., Fischer, P., & Brox, T. (2015). *U-Net: Convolutional Networks for Biomedical Image Segmentation*. https://doi.org/10.48550/ARXIV.1505.04597

Satellite Imaging Corporation. (2022, June). *IKONOS Satellite Imagery, Satellite Specifications | Satellite Imaging Corp*. https://www.satimagingcorp.com/satellite-sensors/ikonos/

Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on Image Data Augmentation for Deep Learning. *Journal of Big Data*, *6*(1), 60. https://doi.org/10.1186/s40537-019-0197-0

Shuvo, F. K., Feng, X., Akaraci, S., & Astell-Burt, T. (2020). Urban green space and health in low and middle-income countries: A critical review. *Urban Forestry & Urban Greening*, *52*, 126662. https://doi.org/10.1016/j.ufug.2020.126662

Simonyan, K., & Zisserman, A. (2014). *Very Deep Convolutional Networks for Large-Scale Image Recognition*. https://doi.org/10.48550/ARXIV.1409.1556

Skår, M., Nordh, H., & Swensen, G. (2018). Green urban cemeteries: More than just parks. *Journal of Urbanism: International Research on Placemaking and Urban Sustainability*, *11*(3), 362–382. https://doi.org/10.1080/17549175.2018.1470104

Smart Dublin. (2020). *Dublin's Open Data Portal*. https://data.smartdublin.ie/

Sperandelli, D. I., Dupas, F. A., & Dias Pons, N. A. (2013). Dynamics of Urban Sprawl, Vacant Land, and Green Spaces on the Metropolitan Fringe of São Paulo, Brazil. *Journal of Urban Planning and Development*, *139*(4), 274–279. https://doi.org/10.1061/(ASCE)UP.1943-5444.0000154

TLV OpenData. (2020). *Home*. https://opendata.tel-aviv.gov.il/en/Pages/home.aspx

Twohig-Bennett, C., & Jones, A. (2018). The health benefits of the great outdoors: A systematic review and meta-analysis of greenspace exposure and health outcomes. *Environmental Research*, *166*, 628–

637. https://doi.org/10.1016/j.envres.2018.06.030

U.S. Geological Survey. (2022, June). *What are the band designations for the Landsat satellites? | U.S. Geological Survey*. https://www.usgs.gov/faqs/what-are-band-designations-landsat-satellites

Wang, S., Wang, M., & Liu, Y. (2021). Access to urban parks: Comparing spatial accessibility measures using three GIS-based approaches. *Computers, Environment and Urban Systems*, *90*, 101713. https://doi.org/10.1016/j.compenvurbsys.2021.101713

World Cover. (2020). *Worldwide land cover mapping*. https://esa-worldcover.org/en

World Health Organization. (2017). *Urban green spaces: A brief for action*.

Xu, Z., Zhou, Y., Wang, S., Wang, L., Li, F., Wang, S., & Wang, Z. (2020). A Novel Intelligent Classification Method for Urban Green Space Based on High-Resolution Remote Sensing Images. *Remote Sensing*, *12*(22), 3845. https://doi.org/10.3390/rs12223845

Yao, X. (2009). Georeferencing, Geocoding. In *International Encyclopedia of Human Geography* (pp. 458–465). Elsevier. https://doi.org/10.1016/B978-008044910-4.00448-X

Yu, C., & Hien, W. N. (2006). Thermal benefits of city parks. *Energy and Buildings*, *38*(2), 105–120. https://doi.org/10.1016/j.enbuild.2005.04.003

Zhao, W., Du, S., Wang, Q., & Emery, W. J. (2017). Contextually guided very-high-resolution imagery classification with semantic segments. *ISPRS Journal of Photogrammetry and Remote Sensing*, *132*, 48–60. https://doi.org/10.1016/j.isprsjprs.2017.08.011

Zylshal, Sulma, S., Yulianto, F., Nugroho, J. T., & Sofan, P. (2016). A support vector machine object based image analysis approach on urban green space extraction using Pleiades-1A imagery. *Modeling Earth Systems and Environment*, *2*(2), 54. https://doi.org/10.1007/s40808-016-0108-8