

Project Resume: CNN to Study Green Areas and Land Use in Cities Over Time

Tyler Brown, Artem Minakow, Elena Putilova, Jödis Strack

Project Description

This project explores the application of Convolutional Neural Networks (CNNs) to study green areas and land use changes in cities over time using satellite and aerial imagery. By leveraging advanced CNN architectures, models capable of identifying and analyzing the evolution of urban green spaces and land use patterns are developed. This approach aids urban planners and environmental researchers in understanding the dynamics of urban growth and environmental sustainability.

For similar studies, refer to:

- Study 1
- Study 2

Project Steps

Part 1: Data Collection and Preliminary Analysis

To generate datasets, we leverage OpenStreetMap with its over 7.5 billion data points (“OpenStreetMap”, 2024), and satellite imagery from ESA’s Sentinel-2 mission (“Sentinel-2”, 2024), to create a comprehensive representation of land use and green spaces for the city of Wuhan, China. The region of interest is defined using a bounding box that covers specific coordinates [30.770277, 30.314773, 114.629974, 113.935089].

Data Collection

We begin by utilizing the OSMnx library to download geometrical data from OpenStreetMap for our labels, following state of the art research (Grinberger et al., 2022; Ludwig et al., 2021). The geometries obtained by OSMnx are then aggregated into a unified polygon representing all green spaces within the bounding box. Next, we generate a grid of 512x512 pixels overlaying the bounding box, where each cell in the grid represents a 10-meter resolution segment of the area. By specification of relevant tags, such as leisure, natural, landuse, waterway, and amenity to include only features related to green spaces—like parks, gardens, forests, and bodies of water. If a cell is marked as any tag, we classify it as either green space (1) or non-green space (0).

For our features, we retrieve satellite images for a time period of Mai to August of 2022 using the Sentinel Hub API and the coordinates specified above. Our features are true colour images with minimal cloud cover

from Sentinel-2 to ensure the dataset’s quality and relevance. The satellite image tiles are stored as numpy arrays and stitched together to form a complete image of the region for each date in the time interval, resulting in a series of images that capture the area’s appearance at different times. As a final step we flip and adjust the matrices of feature and label data to synchronise the two datasets for analysis.

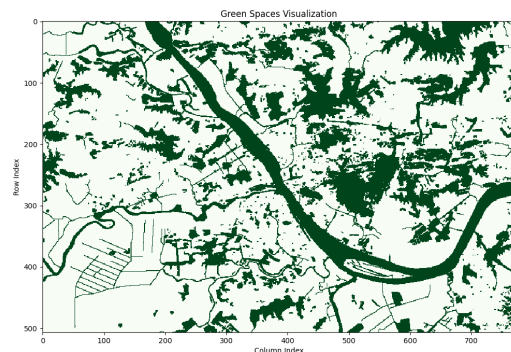


Figure 1: An example mapping of a Sentinel-2 feature and its corresponding OSM label

Slicing Window

To create datasets intended for use as training data in a machine learning model, we segmented the images into smaller patches. This was accomplished using a sliding window technique, which systematically moves across each image, extracting patches of a defined size (512x512 pixels). As the window slides over the images, these patches are captured and stored for further use.

The sliding window’s step size begins at a specified value and increases with each successive image in the dataset, ensuring varying degrees of overlap between patches from different images. This variability enhances the diversity of the training samples. Once the patches are extracted, they are stored in lists, which are then converted into NumPy arrays for efficient processing in the subsequent stages of model training. This process is crucial for preparing a comprehensive and varied dataset that effectively captures different regions of the original images.

Data Cleaning

All satellite images were downloaded directly from the Sentinel Hub API, which is known for providing high-quality, well-maintained data. Yet, images may still

be corrupted due to network interruptions during the download process. To address this, we performed an initial inspection by loading each image to verify its integrity. None of the images we obtained were corrupted. To ensure the validity and robustness of our feature-label-mapping, we used the `bbox_to_dimensions` function from Sentinel Hub, to ensure that all images were generated at a consistent 10 meter resolution. This was crucial for aligning the satellite images with the land classification grid and preventing discrepancies in pixel dimensions. Finally, visual inspections of the merged images confirmed to us that all tiles had been correctly stitched together without visible seams or misalignments (Figure 2).

Preliminary Data Analysis and Visualization

The total area of Wuhan is covered by XX.XX% Urban Green Space, see figure 1.



Figure 2: An example mapping of a Sentinel-2 feature and its corresponding OSM label

Part 2: Deep Learning Model

Build CNN Model

The model implemented for the project goals is a convolutional neural network (CNN) specifically designed for image segmentation tasks. Its architecture follows a U-Net structure, which is composed of an encoder (downsampling path), a bottleneck, and a decoder (upsampling path). This model is particularly effective for tasks where precise localization is important, such as image segmentation.

The encoder consists of two blocks of convolutional layers that progressively downsample the input image while increasing the depth (number of feature channels). Each block in the encoder reduces the spatial dimensions of the input, allowing the model to capture increasingly abstract features.

First Encoder Block starts with two convolutional layers, each using 64 filters with a kernel size of 3x3. Padding is applied to maintain the spatial dimensions after each convolution. After each convolution, a ReLU activation function is applied to introduce non-linearity. A max pooling operation is then applied, reducing the spatial dimensions of the feature map by half (i.e., from 48x48 to 24x24).

Second Encoder Block is similar to the first, but it uses 128 filters for each convolutional layer. The max pooling operation is again applied, further reducing the spatial dimensions (i.e., from 24x24 to 12x12).

The bottleneck serves as the bridge between the encoder and decoder. It captures the most compressed and abstract representation of the input image. This section contains two convolutional layers with 256 filters each and uses a 3x3 kernel size. The padding is maintained, and ReLU activations are applied after each convolution.

The decoder mirrors the encoder's structure but in reverse. It progressively upsamples the feature maps to reconstruct the spatial dimensions of the input image while maintaining the feature information captured during the encoding phase. The key feature of the U-Net decoder is the use of skip connections, which concatenate feature maps from the corresponding encoder layers to retain high-resolution details. First Decoder Block starts with a transposed convolution (also known as a deconvolution) to upsample the feature map. The transposed convolution doubles the spatial dimensions from 12x12 to 24x24 and reduces the depth from 256 to 128 channels. The upsampled feature map is concatenated with the feature map from the second encoder block (skip connection), resulting in a feature map with 256 channels. This is followed by two convolutional layers with 128 filters each and ReLU activations.

Layer (type)	Output Shape	Param #
Conv2d-1	[-1, 64, 48, 48]	1,792
ReLU-2	[-1, 64, 48, 48]	0
Conv2d-3	[-1, 64, 48, 48]	36,928
ReLU-4	[-1, 64, 48, 48]	0
MaxPool2d-5	[-1, 64, 24, 24]	0
Conv2d-6	[-1, 128, 24, 24]	73,856
ReLU-7	[-1, 128, 24, 24]	0
Conv2d-8	[-1, 128, 24, 24]	147,584
ReLU-9	[-1, 128, 24, 24]	0
MaxPool2d-10	[-1, 128, 12, 12]	0
Conv2d-11	[-1, 256, 12, 12]	295,168
ReLU-12	[-1, 256, 12, 12]	0
Conv2d-13	[-1, 256, 12, 12]	590,080
ReLU-14	[-1, 256, 12, 12]	0
ConvTranspose2d-15	[-1, 128, 24, 24]	131,200
Conv2d-16	[-1, 128, 24, 24]	295,040
ReLU-17	[-1, 128, 24, 24]	0
Conv2d-18	[-1, 128, 24, 24]	147,584
ReLU-19	[-1, 128, 24, 24]	0
ConvTranspose2d-20	[-1, 64, 48, 48]	32,832
Conv2d-21	[-1, 64, 48, 48]	73,792
ReLU-22	[-1, 64, 48, 48]	0
Conv2d-23	[-1, 64, 48, 48]	36,928
ReLU-24	[-1, 64, 48, 48]	0
Conv2d-25	[-1, 1, 48, 48]	65
Total params: 1,862,849		
Trainable params: 1,862,849		
Non-trainable params: 0		

Figure 3: The architecture of our CNN model

Similar to the first decoder block, a transposed convolution is used to upsample the feature map from 24x24 to 48x48, reducing the depth to 64 channels. The upsampled feature map is concatenated with the feature map from the first encoder block, resulting in a feature map with 128 channels. Two convolutional layers with 64 filters each and ReLU activations are applied to refine the feature map.

The final layer of the U-Net is a 1x1 convolution that reduces the depth of the feature map to the desired number of output channels. In this case, since the task is binary segmentation, the output is a single-channel feature map where each pixel represents the probability of belonging to the foreground class.

During the forward pass, the input image is first processed by the encoder to extract features at multiple levels of abstraction. These features are then passed through the bottleneck to capture the most condensed information. The decoder then reconstructs the spatial dimensions of the image while utilizing the skip connections to retain high-resolution information from the encoder. The final output is a segmentation map that matches the input image’s spatial dimensions. In total the model consist of 25 layers and 1 862 849 parameters (Figure 3)

Fine-tuning and Hyperparameters

The batch size is set to 10, meaning that the model will process 10 samples at a time before performing a backward pass and updating the weights. The model trained for 70 epochs, meaning it will go through the entire training dataset 70 times. Early stopping triggered at 51 epoch with patience parameter equal 5. By the end of the 51 epochs, both the training loss and validation loss appear to have converged to low values, with minimal fluctuations. This suggests that the model has reached a point where further training might not yield significant improvements (Figure 4).

A learning rate, which controls the step size at each iteration while moving toward a minimum of the loss function, is set to 0.001. The regularization parameter is set to 0.0001. This small value adds a penalty to the loss function proportional to the magnitude of the weights. It helps to keep the model simpler by discouraging the model from fitting to noise in the training data, thereby improving its generalization to unseen data.

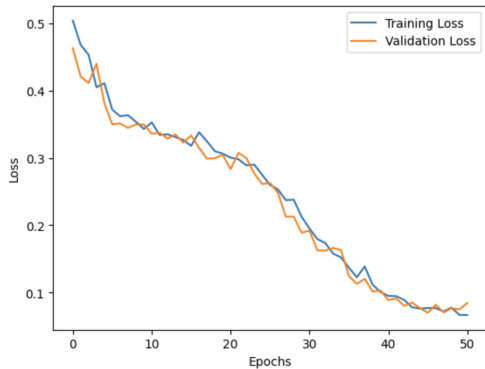


Figure 4: Training and validation loss over the course of 51 epochs

Model Training and Validation

The implemented model shows strong overall performance in classification tasks with test data (Figure 3).

With an accuracy of 0.9626, the model correctly classifies 96.26% of all instances, demonstrating high reliability. The precision of 0.9314 shows that when the model predicts a positive class, it is correct 93.14% of the time, indicating a low rate of false positives. Meanwhile, a recall of 0.9068 reveals that the model successfully identifies 90.68% of all actual positive instances, meaning it is effective at capturing the majority of green zones on the satellite images. The F1 score of 0.9189, which balances precision and recall, underscores the model’s ability to maintain a good equilibrium between correctly identifying positive cases and avoiding false alarms. These metrics collectively suggest that the model performs well in distinguishing between classes, with a robust ability to generalize across different instances.

Part 3: Urban Planning and Environmental Analysis

Urbanization in China impacts millions of people’s lives: moreover, physical and mental health (Gong et al., 2018), economic inequality (Wu and Rao, 2016), and political freedoms (Cartier, 2014) are intimately linked to the intensive and widespread growth of cities. Greenspace, a key component of the built environment, is both a metric of urbanization as well as an outcome of urban development important to supporting human health and well-being.

The literature on urban greenspace (UGS) identifies three important features: quantity, quality, and access. Quantity refers to the total proportion of green space in a given area; while operationalizations of green space vary, they can be “broadly defined as any vegetation found in the urban environment, including parks, open spaces, residential gardens, or street trees” (Kabisch and Haase, 2013). Quality “refers to any attribute that can affect willingness to use and interaction of users with that space,” which may include size, features, and overall condition (Nguyen et al., 2021). Access refers to the topology of greenspace within a city, in particular how easy it is for a resident to reach greenspace within a given time and distance (Barbosa et al., 2007).

With regard to these metrics, the development of UGS in Chinese cities faces several challenges. Wang (2009) argues that Chinese urban planning emphasizes quantity over quality, whereby newly constructed greenspaces are poorly integrated into cities and unevenly distributed between central and peripheral districts. Administrative restrictions are one obstacle; however, the overall approach to greenspace development is also problematic, relying on a “filling-in” strategy that marginally expands existing green spaces instead of developing new ones.

This leads to patchy green space coverage, where UGS areas are small and fragmented. While expansion serves UGS quantity, contiguous green spaces are regarded as higher quality; Sharifi et al. (2021) find that only large, continuous UGS provides significant benefits to perceptions of well-being. The topology of green space is also influenced by demographic characteristics, with older age, lower education, and worse

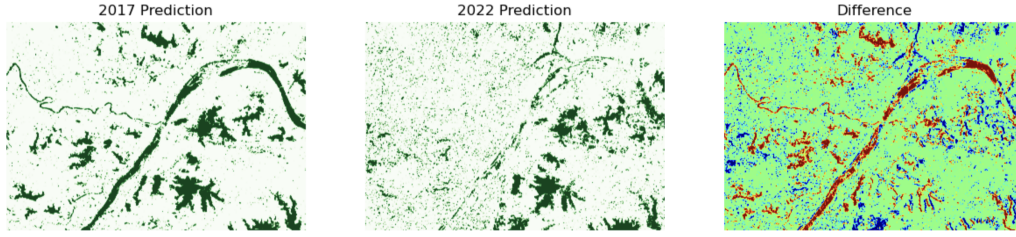


Figure 5: Change in Urban Green Space in Wuhan from 2017 to 2022

housing quality all increasing the fragmentedness and reducing the density of green space (Yang et al., 2022). Nevertheless, fragmentation does not seem to have an impact on overall use, which suggests that there is an inequality in the quality of green space with fragmentation, but not in access (Dong et al., 2020).

Urban Greenspace in Wuhan

Wuhan is an important case among Chinese cities for understanding how urbanization and UGS development interact. As an inland city that faced deindustrialization in the late 20th Century, Wuhan’s contemporary expansion has proceeded rapidly and recently compared to other major cities in China along the coast (Wang et al., 2022). Expansion in the late 2000s and early 2010s was supported by the regional government of the Hubei province, which capitaluted in the 2017 “Modernization, Internationalization, Greenization Plan of Greater Wuhan,” a long-term urban revival plan targeted at improving the city by 2035 (ibid.).

Zhai et al. (2021) uses a support vector machine (SVM) model to analyze changes in land use/land cover (LULC) patterns in Wuhan, finding that between 2000 and 2019, the proportion of built space (i.e. land with structures, roads, etc.) increased 228 percent, while cropland reduced by 12 percent.

They also find that the percentage of “natural habitats” in the study area fluctuates year by year. This suggests that green space has developed alongside built space, but not in one to one complementarity. Their analysis also finds that much of the expansion of built space occurred in the northern suburbs of the city instead of the center, which is likely already saturated.

Empirical observation of UGS development in Wuhan

To empirically evaluate the change in green space in Wuhan, we make descriptive comparisons of green space coverage over time using predicted land use labels from our model. Without the model, prediction over time would not be possible, as the labels are temporally static. However, after training the model on contemporary data, we are able to predict labels from past data as well, which allows us to make fine-grained predictions about changes in greenspace coverage. We take 2017 as our base year, which coincides with the introduction of the Modernization, Internationalization, Greenization Plan. For comparison, we use imagery from 2022,

roughly the time period our labels correspond to (Figure 5).

After predicting labels on the satellite imagery, we take a simple difference of the labels and map “red” to a decrease in the probability of being greenspace and “blue” to an increase. From the model results, we can see that, strikingly, greenspace is predicted to have shrunk considerably over the five-year period, particularly north of the Yangtze river. In the south, we see that the reduction in greenspace is less pronounced, though the greenspace around the southern lakes in the city has decreased slightly.

Limitations

One major modeling obstacle is bodies of water: Wuhan has many lakes and rivers, all of which are labeled identically to land greenspace. Accordingly, the model likely underperforms on land greenspace because it is also trained on largely homogeneous stretches of water. Due to data and computation constraints in the creation and use of our labels, we were not able to separately identify and remove bodies of water from the greenspace classification. Moreover, the albedo (surface reflection) of water varies significantly across our data, which could potentially depend on pollution or time of day. This potentially explains why the rivers are not classified as greenspace in our 2022 samples, as their color changes significantly over time.

Another important concern is that cropland and greenspace may be similarly identified. One important diagnostic that would help us better understand the robustness of our model is the precision on a subset of pixels containing agricultural land and UGS only, which would better quantify how well the model separates urban greenspace from other natural environments. The overall precision and recall scores suggest this is not a large issue; however, this is potentially related to high discriminative power between urban and green-or-crop space and not informative of more subtle distinctions within greenspace classifications. However, the fact that precision (0.93) is only slightly greater than recall (0.91) suggests that the model has good overall predictive power. The F1 score (0.91) is lower than the accuracy (0.96), which is reasonable given that only 24 percent of our study area is covered with greenspace.

References

- Grinberger, A. Y., Minghini, M., Juhász, L., Yeboah, G., & Mooney, P. (2022). OSM Science—The Academic Study of the OpenStreetMap Project, Data, Contributors, Community, and Applications. *ISPRS International Journal of Geo-Information*, 11(4), 230. <https://doi.org/10.3390/ijgi11040230>
- 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>
- OpenStreetMap. (2024, August). Retrieved August 10, 2024, from <https://www.openstreetmap.org/>
- Sentinel-2. (2024, August). Retrieved August 10, 2024, from <https://copernicus.eu/copernicus/sentinel-2>