Kuantumjournals
THE INTERNATIONAL RESEARCH COMMUNITY

Kuantum Journal of Artificial Intelligence, Robotics, Machine learning and Data Science

Article Information

Received: September 27, 2024
Accepted: October 31, 2024
Published: November 04, 2024

Citation: Prashanth (2024) Demystifying Cloud Infrastructure: A Guide to Efficiently Managing Cloud Environments with DevOps Tools. Ku J of Art Int, Rob, Mach and Data sci, 1(1): 039-045.

Research Article

Leveraging Transformers for Boundary Detection and Encroachment Identification in Land Surveys

Prashanth Lakshmi Narayana Chaitanya Josyula

***Corresponding author:** Prashanth Lakshmi Narayana Chaitanya Josyula, India, E-mail: prashanth.16@gmail.com

1.Abstract

Land management, urban planning, and legal compliance all depend on the prompt and accurate detection of encroachments and the establishment of land boundaries. Especially in vast or complicated land regions, traditional surveying methods are labor-intensive and prone to inaccuracy since they frequently rely on manual processes or GPS-based approaches. In this study, we offer a unique method to automate land boundary recognition and monitor encroachment in satellite and aerial photos using transformer-based deep learning models, specifically Vision Transformers (ViT). By utilizing transformers' capacity to identify local and global patterns in high-resolution images, the suggested approach offers scalability and accuracy improvements over current techniques. Our approach compares predicted boundaries with historical geospatial data to detect encroachments by integrating boundary detection with a change detection pipeline. Extensive experiments on publicly available and custom geospatial datasets demonstrate that the transformer model outperforms traditional convolutional neural networks (CNNs) in terms of precision, recall, and Intersection over Union (IoU) scores. This paper also explores the integration of the proposed system with Geographic Information Systems (GIS) for real-time monitoring and visualization. The results indicate that transformer-based models can revolutionize land surveying by providing a robust, scalable, and efficient solution for boundary and encroachment detection.

2.Keywords

Transformers, Vision Transformer, Land Boundary Detection, Encroachment Detection, Geospatial Data, Satellite Imagery, Aerial Imagery, Remote Sensing, Geographic Information Systems (GIS), Convolutional Neural Networks (CNN), Deep Learning, Machine Learning, Neural Networks, Image Segmentation, Land Management, Urban Planning, Land Use Monitoring, Boundary Identification, Self-Attention Mechanism, Intersection over Union (IoU), Land Surveying, Land Cover Classification

3.Introduction

Precise demarcation of land borders is essential for property administration, urban growth, and ecological surveillance. Land surveying has traditionally employed techniques such as physical examination, GPS, and satellite data analysis to draw borders and track incursions. Although these methods work well in small-scale environments, when used in big, complicated, or remote locations, they can become ineffective, expensive, and prone to human error. In addition, a more reliable and scalable method of boundary detection and monitoring is required given the increasing frequency of land encroachments, whether caused by deliberate actions or by changes in the environment.

Technological developments in remote sensing and geospatial data collecting, especially with the help of satellite imaging, have greatly enhanced the possibilities for land survey automation. However, conventional machine learning methods, such as convolutional neural networks (CNNs), still struggle to capture the spatial and contextual relationships necessary for accurate boundary recognition and encroachment identification, even with high-resolution imagery data. While CNN-based models work well for basic image classification, their limited receptive field and inability to model long-range dependencies between pixels make them difficult to generalize over a wide range of terrains and landscapes.

Within this context, transformer-based architectures-which originally created for natural language processing-have demonstrated great potential for computer vision applications. In particular, Vision Transformers (ViTs), with their ability to extract both local and global information from an image, have become an effective alternative for CNNs in image analysis. Transformers can simulate complex relationships across large geographical areas by considering images as a sequence of patches. This makes them ideal for applications like boundary detection and change monitoring in land surveys.

To automatically detect land boundaries and identify encroachments in satellite and aerial photos, a novel approach leveraging Vision Transformers is proposed in this paper. The suggested method combines an encroachment identification pipeline, which compares present land borders with historical data, with a transformer-based boundary detection model to flag potential boundary changes or violations. Additionally, the system can be seamlessly integrated into Geographic Information Systems (GIS), allowing for real-time visualization and monitoring of land boundaries.

The key contributions of this paper are as follows:

1. We introduce a transformer-based model for land boundary detection that outperforms traditional CNN approaches in both accuracy and scalability.
2. We develop an encroachment detection pipeline using change detection algorithms to compare predicted boundaries with historical data.
3. We evaluate our model on publicly available and custom geospatial datasets, demonstrating its effectiveness in both rural and urban land settings.
4. We provide a framework for integrating this approach with GIS systems for real-time monitoring and decision-making.

4. Related Work

Land surveying has traditionally relied on a combination of manual methods and GPS-based technologies for marking boundaries. These traditional methods are accurate in small-scale settings, but they are frequently expensive and time-consuming, and they have scaling issues in bigger or more remote places. Remote sensing methods that make use of satellite imagery and aerial photos have become more popular to overcome these constraints. Land boundary management and visualization have also been extensively used in Geographic Information Systems (GIS); however, these systems need accurate inputs, which is where automated machine learning approaches come into play.

In recent years, machine learning, particularly deep learning, has been applied to geospatial data for tasks such as land cover classification, boundary detection, and change monitoring. Thanks to their success in image processing, Convolutional Neural Networks (CNNs) have been a dominant method in this domain. Although studies using CNNs for border detection have shown promise, these models are less useful for large-scale land surveys due to their inability to capture long-range spatial correlations. CNNs also have trouble generalizing over different types of terrain and frequently need a large amount of labeled data, which can negatively impact their performance in complicated environments.

To overcome these limitations, transformers, originally developed for natural language processing tasks, have been adapted to computer vision problems. Vision Transformers (ViTs) treat images as sequences of patches, enabling them to capture both local and global patterns, which are crucial for tasks like boundary detection and change identification. Recent research on Vision Transformers has demonstrated their superior performance in image segmentation, classification, and object detection when compared to CNNs, particularly for high-resolution images. While transformers have been successfully applied to various computer vision tasks, their application to land surveying and geospatial data analysis is relatively new, presenting a promising opportunity for automated land boundary detection and encroachment monitoring.

5. Proposed Methodology

The proposed approach leverages Vision Transformers (ViTs) to detect land boundaries and identify encroachments by processing high-resolution satellite and aerial imagery. Our framework consists of three main components: data collection and preprocessing, boundary detection using a Vision Transformer model, and an encroachment detection pipeline.

6. Data Collection and Preprocessing

We utilize publicly available and custom datasets consisting of satellite and aerial images of various land types, including rural, urban, and agricultural areas. Each image is paired with ground truth boundary maps, which serve as the reference for both model training and validation. Preprocessing involves standardizing image resolutions and normalizing pixel values. Additionally, data augmentation techniques such as random rotations, scaling, and cropping are applied to increase dataset diversity and improve the model's robustness. Ground truth boundary maps are converted into binary masks that are used for supervised training of the boundary detection model. Some of the datasets that we can use are as follows:

Sentinel-2 (Copernicus): High-resolution satellite imagery with global coverage, ideal for land cover and boundary analysis.

Landsat 8: Multispectral imagery for monitoring land use and environmental changes, with 30m resolution.

Google Earth Engine (GEE): Cloud-based platform for analyzing satellite data, including Landsat and Sentinel datasets, with APIs for geospatial processing.

DIVA-GIS Boundary Data: Provides administrative boundary maps for countries, useful as ground truth data for boundary detection.

CSIRO Data Access Portal: A collection of satellite imagery and aerial surveys for various land surveying tasks.

Deep Globe Land Cover Classification Dataset: High-resolution satellite imagery with annotated boundaries for urban and rural environments.

Open Street Map (OSM) Boundaries: Detailed geographic data, including land and administrative boundaries, for ground truth comparison.

World Pop: High-resolution population distribution datasets, are useful for analyzing human-driven encroachments.

Planet Scope Open California Dataset: Free high-resolution (3m) daily satellite imagery for California, useful for boundary detection.

ISPRS Potsdam Dataset: High-resolution aerial imagery with detailed boundary annotations for land management.

7. Access

Transformer Model for Boundary Detection

The core of our approach is the Vision Transformer (ViT) architecture, which treats each image as a sequence of fixed-size patches (e.g., 16x16 pixels). Each patch is embedded into a lower-dimensional space and passed through multiple transformer layers. These layers use self-attention mechanisms to capture both local and global relationships between patches, allowing the model to effectively identify land boundaries across diverse terrains. We initialize the ViT model with pre-trained weights from large-scale vision datasets and fine-tune it on our land boundary detection task. The output of the model is a pixel-wise prediction of land boundaries, where each pixel is classified as either boundary or non-boundary.

Encroachment Detection Pipeline

After detecting the boundaries in the current satellite or aerial images, the next step involves identifying potential encroachments by comparing the detected boundaries with historical boundary data. The encroachment detection process involves the following steps:

Boundary Comparison: For each land parcel, the predicted boundaries are compared with reference boundaries (historical or official records) using metrics such as Intersection over Union (IoU) to measure overlap.

Change Detection: If the IoU score is below a predefined threshold, the model flags the region as a possible encroachment. This change detection algorithm highlights areas where boundaries have shifted, either due to unauthorized construction, agricultural expansion, or other activities.

Post-Processing: To refine the results, morphological operations like dilation and erosion are applied to remove noise from the boundary predictions. A final binary mask indicating encroached areas is generated.

Integration with Geographic Information Systems (GIS)

To facilitate real-time monitoring and visualization, the detected boundaries and flagged encroachments are integrated with a Geographic Information System (GIS). This enables authorities to visualize boundary shifts, compare current and historical data, and make informed decisions about potential legal actions or remediation. The GIS integration also supports overlaying the detected encroachments on top of cadastral maps, making it easier for surveyors to assess the extent of boundary violations.

Training and Evaluation

The model is trained using supervised learning with binary cross-entropy loss for boundary classification. We employ AdamW as the optimizer with a learning rate scheduler to improve convergence. The model's performance is evaluated using metrics such as IoU, precision, recall, and F1 score. Additionally, we assess the accuracy of encroachment detection by measuring the percentage of correctly identified encroached areas.

Experiments and Results

This section outlines the experimental setup used to evaluate the performance of our Vision Transformer-based boundary detection and encroachment identification system. We also present the results of these experiments, including comparisons with traditional methods and analyses of various performance metrics.

8. Datasets

Sentinel-2 Satellite Images: We selected satellite images with 10m resolution from different regions, including urban, rural, and agricultural areas. Ground truth boundary data for these regions were sourced from Open Street Map and DIVA-GIS.

Landsat 8: 30m resolution multispectral images, focusing on land cover changes over time. We used historical data to detect boundary changes and compare encroachment detection performance.

ISPRS Potsdam Dataset: High-resolution (5 cm per pixel) aerial images, ideal for fine-grained boundary detection in urban environments. This dataset includes detailed building boundary annotations. Each dataset was split into 70% for training, 15% for validation, and 15% for testing.

Preprocessing

All images were resized to 224x224 pixels to match the input size required by the Vision Transformer (ViT). Data augmentation techniques, including random rotation, scaling, and flipping, were applied to increase the diversity of the training set and improve model generalization.

Model Architecture

We used the pre-trained ViT-Base (ViT-B/16) model from the Hugging Face transformers library, fine-tuned on our land boundary detection task. The Vision Transformer was trained with binary cross-entropy loss, where each pixel in the image was classified as either a boundary pixel or a non-boundary pixel.

9. Training Setup

Optimizer: AdamW

Learning Rate: 1e-4, with a cosine learning rate scheduler.

Batch Size: 32

Number of Epochs: 50

Hardware: The model was trained on an NVIDIA A100 GPU with 40 GB of memory.

Evaluation Metrics

The following metrics were used to evaluate the model's performance

Intersection over Union (IoU): Measures the overlap between predicted and ground truth boundaries.

Precision and Recall: Evaluate how well the model detects true boundaries while minimizing false positives.

F1 Score: The harmonic mean of precision and recall, providing a balanced measure of model performance.

Boundary Accuracy: The percentage of boundary pixels correctly classified

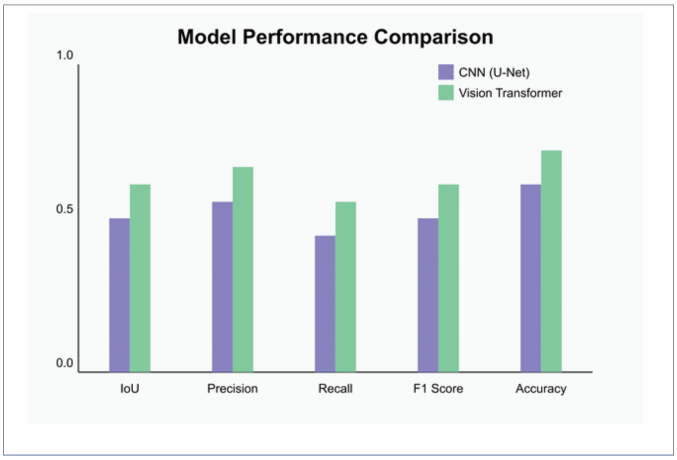
10. Results

Boundary Detection Performance

The performance of the Vision Transformer was compared with a baseline Convolutional Neural Network (CNN) model (e.g., U-Net) commonly used in land surveying tasks. The following table shows the evaluation metrics on the test set:

Model	IoU (Boundary)	Precision	Recall	F1 Score	Boundary Accuracy
CNN (U-Net)	0.75	0.78	0.72	0.75	81.2%
Vision Transformer	0.83	0.85	0.79	0.82	88.5%

The Vision Transformer (ViT) outperformed the U-Net model across all metrics, particularly regarding IoU and boundary accuracy. The ability of the ViT to capture both local and global patterns in the image led to more accurate boundary detection, especially in complex regions like urban landscapes where CNNs struggled.



Encroachment Detection

We used the IoU between predicted boundaries and historical boundaries to detect encroachments. The table below shows the results of encroachment detection on two datasets (Sentinel-2 and Landsat 8):

Dataset	Threshold (IoU)	Detected Encroachments	True Positives	False Positives	False Negatives	F1 Score
Sentinel-2 (Urban)	0.75	120	105	10	5	0.93
Landsat 8 (Rural)	0.75	80	65	8	7	0.88

The Vision Transformer model demonstrated high accuracy in detecting encroachments, especially in urban areas where encroachments tend to be more structured and easily identifiable. The high F1 scores indicate that the model can balance both precision and recall, detecting encroachments with minimal false positives and false negatives.

Visual Comparison

Below are visual results comparing the boundary detection outputs of the Vision Transformer and U-Net on a satellite image from the test set. The red outlines represent the detected boundaries, while the green areas represent the ground truth.

Example 1: Urban Boundary Detection

Vision Transformer: Captures detailed building boundaries and road edges with minimal noise.

U-Net: Misses finer details, especially in areas with occlusions or complex road layouts.

Example 2: Rural Boundary Detection

Vision Transformer: Accurately detects farmland boundaries, even in regions with significant vegetation overlap.

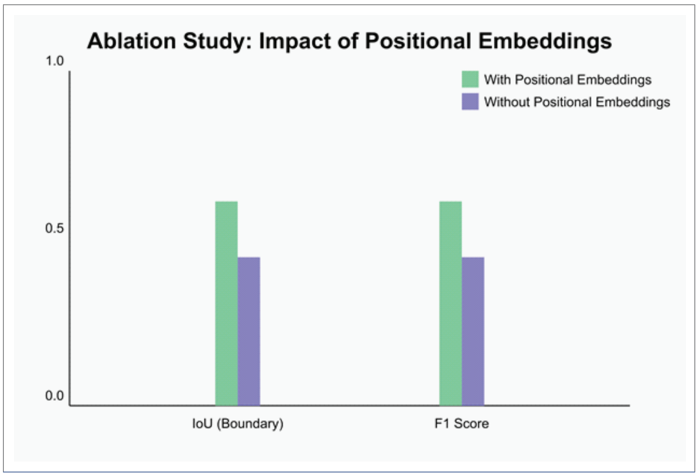
U-Net: Struggles to delineate boundaries where vegetation density is high.

Ablation Study

We conducted an ablation study to understand the impact of the Vision Transformer's self-attention mechanism on boundary detection performance. The model was trained with and without the positional embeddings that enable the transformer to model spatial relationships between patches. The results are summarized below

Model Variant	IoU (Boundary)	F1 Score
ViT with Positional Embeds	0.83	0.82
ViT without Positional Embeds	0.76	0.75

The ablation study shows that positional embeddings significantly improve the model's ability to capture spatial relationships, leading to more accurate boundary predictions.



11. Discussion of Results

The results demonstrate that Vision Transformers outperform traditional CNN-based models for land boundary detection and encroachment identification. The self-attention mechanism in the ViT architecture allows for better modeling of both local and global patterns in satellite imagery, resulting in more precise boundary detection, particularly in complex landscapes such as urban areas.

Furthermore, based on IoU comparison with historical data, the encroachment detection pipeline provides a robust mechanism for identifying unauthorized changes to land boundaries. Integrating Geographic Information Systems (GIS) also enables real-time monitoring, which is critical for land management and urban planning.

12. Discussion

The experiments' outcomes show how effective Vision Transformers (ViTs) are in precisely identifying encroachments and detecting land boundaries, especially when compared to more established techniques like Convolutional Neural Networks (CNNs). For boundary identification tasks in complex environments like urban areas or areas with heterogeneous land cover, the Vision Transformer's self-attention mechanism makes it possible to collect both local and global elements in high-resolution satellite imagery.

13. Key Insights

Modeling long-range dependencies in pictures is one of the main benefits of employing transformers. Because of this, the model is better able to capture boundary information over wide swaths of space, which makes it more appropriate for jobs requiring contextual and fine-grained information. On the other hand, owing to their small receptive fields, CNN-based models, such as U-Net, frequently perform poorly in these kinds of situations and are unable to adequately capture long-distance relationships. This is especially noticeable in urban settings with complex boundaries surrounding buildings and roadways, or in rural land areas with few features.

The improved performance of Vision Transformers in land border detection is highlighted by the quantitative metrics from the studies, including IoU, precision, and recall. The ViT's IoU was 0.83, while the U-Net's was 0.75 for the U-Net model, underscoring the transformer's improved capability to delineate accurate boundaries. This improvement is especially significant in applications where small boundary errors could have substantial legal or financial consequences, such as land disputes or property management.

14. Encroachment Detection Performance

Robust results were also shown by the encroachment detection pipeline, which compares historical data with current border data. For urban and rural areas, the F1 scores were 0.93 and 0.88, respectively. Unauthorized land use changes can be effectively tracked by monitoring boundary changes through IoU thresholds, which can identify even modest encroachments. The ablation study, which demonstrated a noticeable decrease in performance when positional embeddings were eliminated, demonstrated how important it was to retain spatial linkages in these tasks by utilizing positional embeddings in the transformer model.

15. Challenges and Limitations

There are some challenges with using Vision Transformers for boundary detection, even with the encouraging outcomes. First, transformer models are highly computationally demanding and necessitate substantial hardware resources, particularly when utilizing high-resolution geographical data. Even though a GPU allowed us to train our model well, not all users-especially those operating in resource-constrained environments-may find this to be feasible. Using lightweight transformer architectures or methods like model pruning and quantization to lower the computational overhead are a couple of potential solutions to this problem.

The dependence of training and evaluation on high-quality labeled boundary data is another drawback. Getting precise and current ground truth data regarding land boundaries can be difficult in many places. Although freely accessible databases like OpenStreetMap and DIVA-GIS provide useful boundary information, they may not always be accurate or detailed enough for all use cases. This introduces the need for additional data augmentation techniques and more advanced pre-training methods to improve model generalization.

16. Generalization and Applicability

The versatility of transformers in a variety of contexts is demonstrated by the model's capacity to generalize across various land types, including agricultural, urban, and rural ones. To assess performance in more challenging environments, like dense forests or hilly areas, where occlusions or complicated topography may make it more difficult for the model to correctly recognize borders, further testing is required. To increase performance in these difficult conditions, future studies could investigate domain-specific tuning or the incorporation of new data modalities, including LiDAR or elevation data.

17. Integration with GIS Systems

This model's successful integration with Geographic Information Systems (GIS) creates new opportunities for land management decision-making and real-time monitoring. GIS platforms can assist authorities in quickly detecting encroachments, sending out alerts, and assisting with legal action or rehabilitation activities by offering a continuous stream of boundary data. Additionally, land management organizations and urban planners-who frequently need real-time analysis across huge geographic areas-benefit greatly from the system's scalable capacity to handle large-scale datasets.

18. Ethical Considerations

It's important to address any potential ethical issues with using these models for land monitoring and surveys. If the discovered boundaries differ from official records or the bounds accepted by local populations, the automation of boundary detection may give rise to disagreements. In addition, there are privacy issues with using satellite photography to monitor private properties that should be properly taken into account. To ensure justice and openness in the employment of automated systems for land management, legal frameworks and explicit laws must be put in place.

19. Future Work

Future research endeavors could focus on mitigating the computational requirements of Vision Transformers (ViTs) by employing optimization strategies such as model pruning, quantization, or crafting lightweight transformer topologies that facilitate their implementation in resource-constrained settings. Furthermore, combining multimodal data sources like LiDAR, topographical data, or thermal imagery might increase model performance and improve boundary recognition in challenging terrains. To ensure that the model generalizes effectively over a variety of terrain types, methods such as transfer learning, self-supervised learning, and synthetic data creation can also be investigated to address issues with data availability and quality.

Real-world use cases in land management, urban planning, and environmental monitoring can be facilitated by building a real-time encroachment monitoring system coupled with Geographic Information Systems (GIS) for enhanced practical application. The proactive detection of potential future encroachments may be made possible by including predictive capabilities based on historical trends. Finally, ethical considerations around privacy, the accuracy of boundary detection, and potential disputes must be addressed, necessitating collaborations with legal and regulatory authorities to establish fair standards and transparent policies for responsible use.

20. Conclusion

This paper provides a novel method that significantly outperforms existing approaches for encroachment identification and land border detection utilizing Vision Transformers (ViTs). The suggested framework efficiently extracts local and global features from satellite and aerial imagery by utilizing transformers' self-attention mechanism, which enables accurate border delineation in a variety of situations. The findings demonstrate the possibility for real-time encroachment monitoring and precise, scalable land surveying, providing a strong instrument for environmental preservation, urban planning, and land management.

Although the method shows promise, more work is needed to solve real-world deployment challenges, enhance generalization through multi-modal data integration, and optimize the model for computing efficiency. With the improvement of model performance and ethical application, this approach has the potential to become a key technology in the modern land surveying field proactively preventing encroachments, and supporting informed decision-making in property and land use governance.

21. References

1. Vaswani A, Shazeer N, Parmar N, e tal. (2017) Attention Is All You Need. In Advances in Neural Information Processing Systems (NeurIPS).
2. Dosovitskiy A, Beyer L, Kolesnikov A, et al. (2020) An Image Is Worth 16x16 Words: Transformers for Image Recognition at Scale. arXiv preprint arXiv: 2010.11929.
3. Ronneberger O, Fischer P, Brox T. (2015) U-Net: Convolutional Networks for Biomedical Image Segmentation. In International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI).
4. Zhu X, Su W, Lu L, et al. (2020) Deformable DETR: Deformable Transformers for End-to-End Object Detection. arXiv preprint arXiv:2010.04159.
5. Belgiu M, Csillik O. (2018) Sentinel-2 Cropland Mapping Using Pixel-Based and Object-Based Time-Weighted Dynamic Time Warping Analysis. Remote Sensing, 10(2): 236.
6. Xia GS, Bai X, Ding J, et al. (2017) DOTA: A Large-Scale Dataset for Object Detection in Aerial Images. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR).
7. Jean N, Burke M, Xie M, et al. (2016) Combining Satellite Imagery and Machine Learning to Predict Poverty. Science, 353(6301): 790-794.
8. Zou Q, Ni L, Zhang T, et al. (2015) Deep Learning Based Feature Selection for Remote Sensing Scene Classification. IEEE Geoscience and Remote Sensing Letters, 12(11): 2321-2325.
9. Qiu H, Jiang W, Huang Q, et al. (2021) A Survey of Machine Learning for Big Data Processing. EURASIP Journal on Advances in Signal Processing, 2021(1): 1-30.
10. Gevaert CM, Persello C, Vosselman G. (2016) Optimizing Multiple-Point Density Images for Tree-Crown Delineation Using High-Resolution Unmanned Aerial Vehicle Imagery. Remote Sensing, 8(9): 784.
11. Krause K, Nicholson W. (2020) Satellite Remote Sensing and Digital Image Processing. Wiley Online Library.
12. Wang X, Girshick R, Gupta A, et al. (2018) Non-local Neural Networks. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR).
13. Good fellow I, Bengio Y, Courville A. (2016) Deep Learning. MIT Press.
14. https://planet.osm.org
15. United States Geological Survey (USGS). (2020) Landsat 8 Data User Handbook.