# URBAN TYPOLOGY AND AMENITY CLASSIFICATION OF SATELLITE IMAGERY USING DEEP CONVOLUTIONAL NEURAL NETWORKS

Hamdi Kerem Kucukengin [1,2,†]

[1] Northwestern University

Master of Science in Data Science Program

[2] Github Repository:

https://github.com/hamodikk

† Address to which correspondence should be addressed:

kkucukengin@gmail.com

# Abstract

Understanding the spatial structure and amenity composition of urban neighborhoods is essential for supporting data-driven urban planning, land use analysis, and livability assessment. This project investigates the application of convolutional neural networks (CNNs) to classify neighborhood typology and detect the presence of public amenities using high-resolution satellite imagery. Using Google Earth Engine to extract Sentinel-2 image tiles and OpenStreetMap data to annotate typology and amenity labels, we construct a geospatial dataset covering diverse urban environments. The classification task includes both multiclass typology prediction (e.g. residential, industrial, mixed-use) and multi-label detection of features such as parks, green spaces, and water bodies. This interim report focuses on data acquisition, preprocessing, and automated labeling, establishing the foundation for supervised image classification using CNNs. By combining remote sensing data with open geospatial annotations, this work aims to enable scalable urban structure analysis from overhead imagery alone.

**Keywords:** Remote sensing, convolutional neural networks, urban typology, satellite imagery, Google Earth Engine, OpenStreetMap, land use classification, geospatial data, image labeling, amenity detection

# Contents

# Introduction and Problem Statement

Urban environments can be shaped by population growth, infrastructure development, land use policy, and socio-economic factors. Understanding the structural composition of cities is a fundamental challenge to urban planning, infrastructure design, and sustainability assessment. Traditional approaches to analyzing urban environments often rely on manually collected zoning maps, land use records, and survey data, which can be costly and inconsistent across regions.

Recent advances in satellite imaging and machine learning offer an alternative, scalable approach for understanding urban spaces. High-resolution satellite imagery provides a consistent, global view of our environment, while convolutional neural networks (CNNs) extract semantic features from image data. Combined, these developments create new opportunities for urban typology classifications such as residential, industrial, or mixed-use areas and detecting the presence of critical amenities like parks, cultural landmarks, and green spaces directly from visual data.

This project investigates the use of CNNs to perform typology classification and amenity detection from overhead imagery. The core objective is to build a supervised image classification pipeline that operates on tiles extracted from Sentinel-2 imagery and labeled using open geographic data from OpenStreetMap. Each tile is annotated with both a single urban typology class and multiple amenity presence flags, enabling a multi-task learning setup. The pipeline is designed to be modular and scalable, with potential applications in regional planning, land use monitoring, and livability assessment.

The current progress of the project focuses on building a geospatial dataset by extracting tiles from Google Earth Engine and automatically assigning labels through spatial intersection with vector data from OpenStreetMap. Future progress will involve training CNN-based classifiers, evaluating model performance across various urban contexts, and exploring the use of unsupervised pretraining to improve feature extraction from unlabeled imagery.

# Literature Review

The application of convolutional neural networks (CNNs) to remote sensing data has gained significant traction in the recent years, driven by advances in both computer vision and the increasing availability of satellite imagery. Foundational work in deep learning by Goodfellow, Bengio and Courville (2016) established that CNNs are a powerful architecture for image classification and segmentation tasks, with later refinements focused on transfer learning and residual networks.

In the context of land use and urban structure classification, well-established datasets such as EuroSAT have demonstrated the viability of CNNs for distinguishing urban typologies from Sentinel-2 imagery. Helber et al. (2019) introduced EuroSAT, a labeled dataset of Sentinel-2 tiles for land cover classification across ten classes, including industrial and residential categories. Their results highlight the potential of multispectral imagery and standard CNNs such as ResNet-50 to perform high-accuracy classification on geospatial tasks.

More recently, researchers have explored self-supervised learning approaches to address the challenge of limited labeled data in remote sensing. Ayush et al. (2021) introduced GEOGRAPHY-AWARE contrastive learning (GeoCLR), demonstrating improved performance in land cover classification using unlabeled satellite image patches. This method suggests that self-supervised pretraining could substantially benefit tasks like urban amenity detection where labeled data is sparse or noisy.

In practical terms, this thesis builds on the publicly available Sentinel-2 dataset, which offers a 10-meter resolution of multispectral imagery with global coverage and regular updates. The Google Earth Engine platform facilitates scalable access to these images, and detailed documentation from the European Space Agency (2023) and Earth Engine community resources make it feasible to incorporate this dataset into machine learning workflows for urban analysis. We believe that this work extends the prior work by introducing multi-label amenity detection and automated spatial labeling across cities.

# Data

This project will use two primary sources of data: high-resolution satellite imagery from the Sentinel-2 satellite and geospatial features data from the OpenStreetMap (OSM). Together, these datasets will provide both the visual and the semantic information necessary to support supervised learning for neighborhood classification and amenity detection.

Sentinel-2 is a European Space Agency (ESA) operated satellite that can provide global coverage at 10-20 meter spatial resolution across 13 spectral bands. This project will be using the red, green and blue bands of the level-2A surface reflectance product, accessing it through Google Earth Engine (GEE), which provides preprocessed images with atmospheric correction such as cloud filter. The dataset will be accessed through GEE Python API and geemap interface to generate tiles in the ROI.

We will retrieve the typology and amenity labels by accessing the vector data from OpenStreetMap's OSMnx library. The dominant typology for each tile will be retrieved from

landuse related tags, whereas various other categories will be used for amenity labels. Labeling will be performed by intersecting each tile with the relevant landuse and amenity tags.

For purposes of this project, we will be focusing on the Chicago metropolitan area, though this can be scaled up based on the success and performance of our model. Although currently we focus only on geospatial data, future studies could include time-series components to research temporal aspects such as change detection or urban growth analysis over time.

# Methods

This project will follow a multi-stage pipeline that combines geospatial data extraction, automated labeling, and deep learning for image classification using a custom CNN architecture. The workflow can be broadly divided into four key components: tile generation, label assignment via OSM, image preprocessing, and CNN-based modeling.

## Image Tile Generation

Using the Earth Engine Python API and Sentinel-2 imagery, we generated over 4,200 image tiles across the Chicago metropolitan area. Each tile covers an area of 512x512 meters and is exported as an RGB GeoTIFF with a resolution of 10 meters per pixel—approximately 48x48 pixels after center cropping during preprocessing. To ensure spatial consistency, tiles are generated on a regular grid and filtered to only include those fully contained within a defined bounding box. This helps avoid edge artifacts and partial overlays at the periphery of the region of interest. Although tile spacing produces slight overlaps, efforts were made to constrain centers within a 'safe zone' of the bounding box. All tiles are exported locally to avoid exceeding export queue limits to Google Drive.

## Automated Labeling via OpenStreetMap

Each image tile is assigned two sets of labels by spatially intersecting its geometry with OSM vector data:

- A single-label typology class, which is derived from the land use type intersecting the tile.
- A multi-label amenity vector, indicating presence of any of four selected categories: park, school, museum, and bodies of water.

The vector data is obtained through osmnx, filtered by bounding box and tag category, and then clipped to the region covered by the tiles. Tiles are intersected with the filtered features, and

those containing relevant tags are labeled accordingly. All resulting annotations are stored in a structured CSV file called tile_labels.csv to be used for downstream modeling.

## Image Preprocessing and Dataset Structure

Image tiles are loaded as 3-channel float tensors and normalized. To standardize dimensions, all tiles are center cropped to 48x48 pixels, as initial training produced errors due to inconsistent data dimensions. A custom PyTorch class is used to load both the images and the corresponding labels to enable joint training for both typology and amenity classification. The dataset is split 80/20 into training and validation sets.

## CNN Model Architecture

The CNN model was implemented from scratch using PyTorch (Figure 1). It includes three convolutional layers with ReLU activations and max-pooling operations, followed by a shared fully connected layer that branches into two output heads:

- A softmax classification head for typology classification (20 categories)
- A sigmoid head for binary multi-label amenity detection (4 labeld)

The total number of learnable parameters is moderate. The model was trained for 10 epochs with Adam optimizer and a batch size of 32. The loss function is a combined sum of cross-entropy loss for typology classification and a binary cross-entropy for amenity detection. Evaluation of the model includes validation loss, typology classification accuracy, and amenity ROC AUC scores.

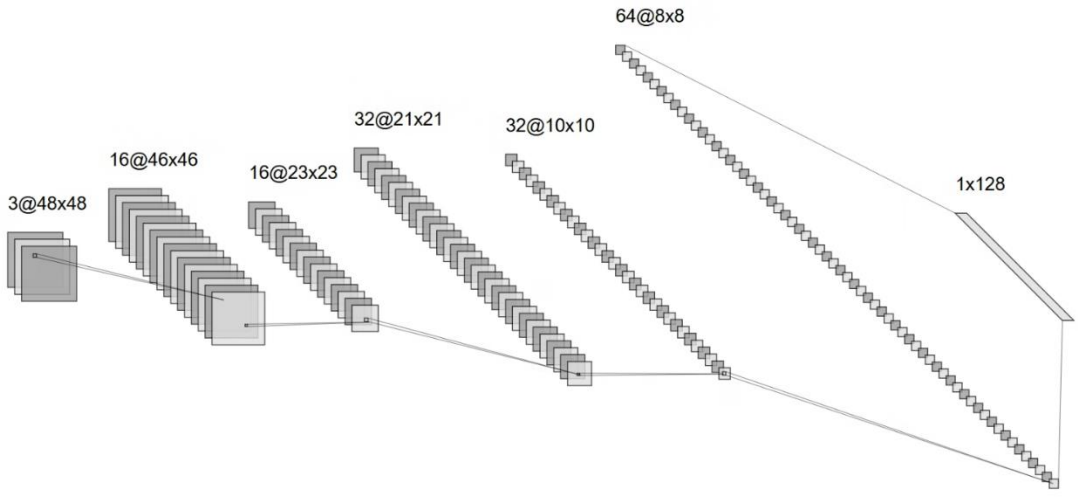


**Figure 1.** ***Architecture of the custom convolutional neural network (CNN) used in this study. The model includes three convolutional layers with ReLU and max pooling, followed by task-specific fully connected layers for typology and amenity classification.***

# Results

This section presents the performance of the custom convolutional neural network (CNN) model trained to classify urban typology and detect amenity presence from Sentinel-2 satellite imagery. Results are organized into two parts: typology classification accuracy and confusion analysis, and amenity detection performance via multi-label evaluation.

## Typology classification

The typology classification task was treated as a single-label prediction problem across 20 distinct land-use categories derived from OpenStreetMap. After 10 training epochs, the model achieved a final validation accuracy of 55.6%. However, the classification report reveals that this accuracy is largely driven by a few dominant classes.

As shown in the confusion matrix (Figure 2) and the classification report (Appendix A), the model predicted the residential class with high recall (0.96) and moderate precision (0.49), making it the most successful prediction across the validation set. The unknown class also performed well, with a near-perfect precision (0.95) and recall (0.99), suggesting the model reliably recognizes when a tile lacks identifiable land use.

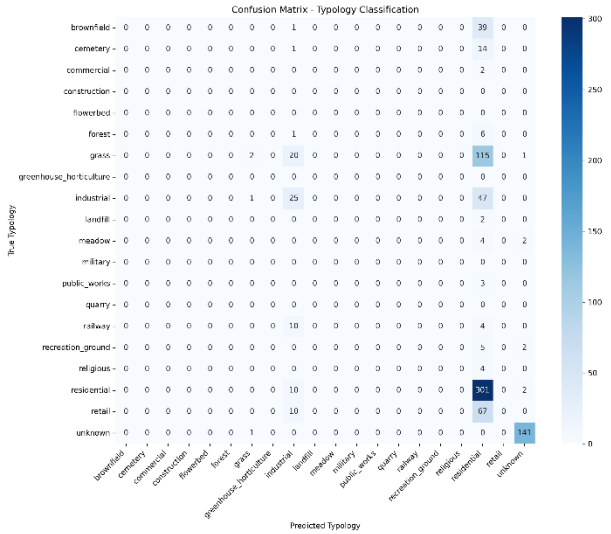


**Figure 2.** ***Confusion matrix showing model predictions for typology classification. Most correct predictions occur for the 'residential' and 'unknown' classes, while rare classes are frequently misclassified.***

Most other typology categories—including public_works and railway—were completely missed, while others such as industrial showed partial learning, also reflected in zero precision and recall. This outcome likely stems from strong class imbalance in the dataset, visual ambiguity of rare typologies in satellite imagery, and the limited representational capacity of the baseline CNN model.

While a few categories such as industrial showed partial learning (precision = 0.32, recall = 0.34), the overall macro-averaged F1-score was low (0.10), underscoring the difficulty of fine-grained typology classification without more extensive model tuning or data balancing.

## Amenity Detection

The amenity head of the model was evaluated as a multi-label classification task. Performance was measured using receiver operating characteristic (ROC) curves and area under the curve (AUC) scores for each amenity tag.

As shown in Figure 3, the model performs particularly well on has_schools (AUC = 0.89) and has_water (AUC = 0.82), with slightly weaker performance on has_park (AUC = 0.75) and has_museum (AUC = 0.69). These scores reflect the relative clarity and consistency of visual cues associated with each amenity type. For example, schools often have distinct shapes and sizes, while museums may be harder to distinguish visually from other buildings.

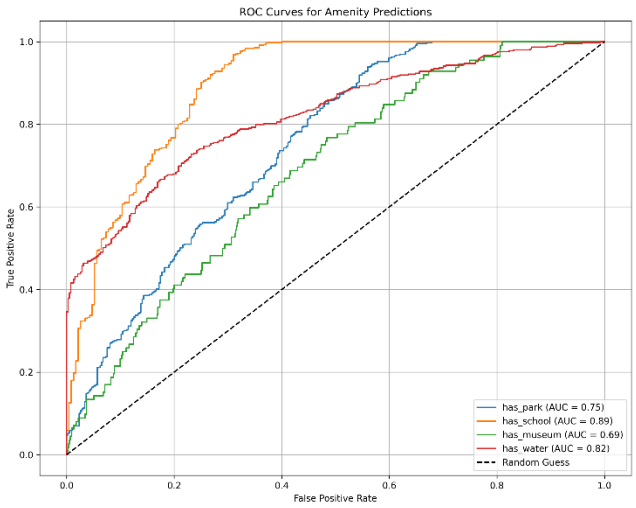


**Figure 3.** ***ROC curves for multi-label amenity detection, showing model performance across four classes: school, water, park, and museum. AUC scores range from 0.69 to 0.89.***

These results indicate that a moderately deep CNN is capable of learning meaningful spatial and textual cues from Sentinel-2 tiles, even in the presence of noisy and incomplete OpenStreetMap labels.

# Discussion

The results of this study demonstrate both the potential and challenges of using convolutional neural networks (CNNs) to classify urban typology and amenity presence from satellite imagery. While the model achieved promising performance in certain categories, the analysis also highlighted key limitations related to data representation, class imbalance, and model complexity.

## Typology Prediction

The typology classification task achieved a validation accuracy of approximately 55.6%. However, a closer look at the confusion matrix and classification report reveals that this performance was heavily influenced by the dominance of a few easily recognizable categories. Specifically, the residential class and the unknown label accounted for the majority of the correctly

predicted samples. Their visual features—such as high-density building patterns or the absence of distinguishable land use—are likely more consistent and detectable by the CNN model.

In contrast, rare classes such as railway, public_works, and retail were not predicted with meaningful accuracy. This is attributed to a combination of low representation in the training set as well as the difficulty of distinguishing these categories from high-resolution imagery without additional context. The model's tendency to default to residential or unknown also underscores the need for class balancing or augmentation strategies to improve learning for less frequent typologies.

## Amenity Detection

The amenity prediction task, framed as a multi-label classification problem, achieved a more consistent and promising result. The AUC score for the four categories ranged from 0.69 to 0.89, suggesting that the model was generally able to detect the visual presence of these urban features. This finding is consistent with prior literature, where visual cues such as vegetation patterns, spatial layouts, or roof colors have been successfully used to identify urban function.

Unlike the typology task, the amenity labels were simpler binary indicators and often represented visually distinct patterns. This supports the hypothesis that CNNs are particularly well-suited for binary visual detection tasks, even when typological classifications are more ambiguous.

## Limitations and Trade-Offs

Several limitations became evident during the development of this pipeline. First, the training dataset was created based on OpenStreetMap features and Sentinel-2 imagery, both of which introduce potential inaccuracies. For example, not all urban features are mapped in OSM, and Sentinel-2's 10-meter resolution can blur finer land use boundaries, especially in dense or mixed zones.

Second, the choice to train a relatively simple CNN architecture (with 3 convolutional layers) was deliberate to maintain interpretability and keep the runtime reasonable, but it likely constrained the model's capacity to extract deeper semantic patterns. Training time was also a practical challenge, with each 10-epoch run requiring over 45 minutes on a standard CPU setup, limiting experimentation time.

Finally, the use of square tiles with a fixed area coverage, while consistent, meant that urban geometry might not always align well within the tiles. For example, OSM could assign a class for a tile based on the partial match of a structure with a tile, but the CNN might not have enough information to train or classify the existence of such structure within the tile. Future work

could explore adaptive tiling, additional data sources, tile transformations, or self-supervised pretraining to address this.

# Conclusions

This thesis presents a complete pipeline for urban typology classification and amenity presence detection using satellite imagery and convolutional neural networks. Leveraging high-resolution Sentinel-2 data, OpenStreetMap annotations, and a custom dataset covering the Chicago metropolitan area, the project demonstrates the feasibility of automating urban classification tasks with deep learning.

The CNN model achieved a typology classification accuracy of 55.6%, with strong performance in dominant categories like residential and unknown, but limited success in rarer or visually ambiguous classes. In contrast, the amenity detection task produced a more balanced and robust result, with ROC-AUC scores up to 0.89 for some features, demonstrating the model's ability to extract and respond to visual cues tied to parks, schools, museums, and water bodies.

From data acquisition and preprocessing to model evaluation and visualization, each component of the pipeline was carefully developed to support extensibility, accuracy and interpretability. Notably, the use of publicly available data sources and the creation of a tile-based dataset tailored to an urban region show how this approach can scale across cities and regions worldwide.

While the results reveal areas for further refinement—including addressing data imbalance and expanding model complexity, this work provides a strong foundation for future research at the intersection of computer vision and urban studies. The methodology and findings outlined here can support applications in urban planning, infrastructure monitoring, and livability analysis, offering both a technical and conceptual roadmap for city-scale analysis using deep learning.

# Directions for Future Work

This study establishes a baseline for automated urban typography and amenity classification using CNNs and remote sensing data. Several avenues remain for future development and refinement:

- **Model Architecture Improvements**

  The current CNN model uses a compact architecture to balance interpretability and training time. Future work could incorporate more expressive architectures similar to

ResNet, EfficientNet, or transformer-based vision models to improve feature extraction, particularly for visually subtle or underrepresented typologies.

- **Self-Supervised Pretraining**

  Implementing self-supervised learning using unlabeled image tiles could improve representation learning prior to fine-tuning on labeled data. This could be beneficial for the imbalanced or sparse sampling of typology classification task.

- **Expanded Label Set and Region Coverage**

  The current label set focuses on four amenity types and 20 typology classes extracted from OSM data. Future versions could expand to include more nuanced features like transit access, zoning categories, etc. The method can also be scaled to other cities or applied globally to explore cross-regional patterns.

- **Time Series and Change Detection**

  While this project uses a single time slice, satellite imagery from previous years could be incorporated to study urban change over time. Techniques such as Siamese networks or hybrid networks utilizing CNNs and RNNs can enable temporal analysis for urban sprawl, green space evolution or infrastructure development.

- **Livability Scoring and Planning Tools**

  A composite livability index derived from typology and amenity predictions could help urban planners assess spatial equity and quality-of-life indicators. Pairing model outputs with interactive dashboards using Folium could transform this research into a city planning support tool.

## Acknowledgements

I would like to express my gratitude to Dr. Alianna Maren for her continued support, insightful feedback, and for fostering an engaging learning environment. I would also like to thank the independent researchers for their contributions to the corpus. Their input played a key role in enabling the work accomplished here.

## Data & Code Availability

Data and the code for this work can be found in the github repository here.

# References


- Ayush, Kumar, Burak Uzkent, Chenlin Meng, Kumar Tanmay, Marshall Burke, David Lobell, and Stefano Ermon. "Geography-aware self-supervised learning." In Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 10181-10190. 2021.
- Goodfellow, Ian, Yoshuo Bengio, and Aaron Courville. Deep Learning. Cambridge, MA: MIT Press, 2016.
- Helber, Patrick, Benjamin Bischke, Andreas Dengel, and Damian Borth. "Eurosat: A novel dataset and deep learning benchmark for land use and land cover classification." IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 12, no. 7 (2019): 2217-2226.
- European Space Agency. “Sentinel-2 User Guide.” Last updated 2023. https://sentiwiki.copernicus.eu/web/s2-applications

# Appendix A

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| brownfield | 0 | 0 | 0 | 40 |
| cemetery | 0 | 0 | 0 | 15 |
| commercial | 0 | 0 | 0 | 2 |
| construction | 0 | 0 | 0 | 0 |
| flowerbed | 0 | 0 | 0 | 0 |
| forest | 0 | 0 | 0 | 7 |
| grass | 0.5 | 0.01 | 0.03 | 138 |
| greenhouse_horticulture | 0 | 0 | 0 | 0 |
| industrial | 0.32 | 0.34 | 0.33 | 73 |
| landfill | 0 | 0 | 0 | 2 |
| meadow | 0 | 0 | 0 | 6 |
| military | 0 | 0 | 0 | 0 |
| public_works | 0 | 0 | 0 | 3 |
| quarry | 0 | 0 | 0 | 0 |
| railway | 0 | 0 | 0 | 14 |
| recreation_ground | 0 | 0 | 0 | 7 |
| religious | 0 | 0 | 0 | 4 |
| residential | 0.49 | 0.96 | 0.65 | 313 |
| retail | 0 | 0 | 0 | 77 |
| unknown | 0.95 | 0.99 | 0.97 | 142 |
| | | | | |
| accuracy | | | 0.56 | 843 |
| macro avg | 0.11 | 0.12 | 0.1 | 843 |
| weighted avg | 0.45 | 0.56 | 0.44 | 843 |

# Appendix B

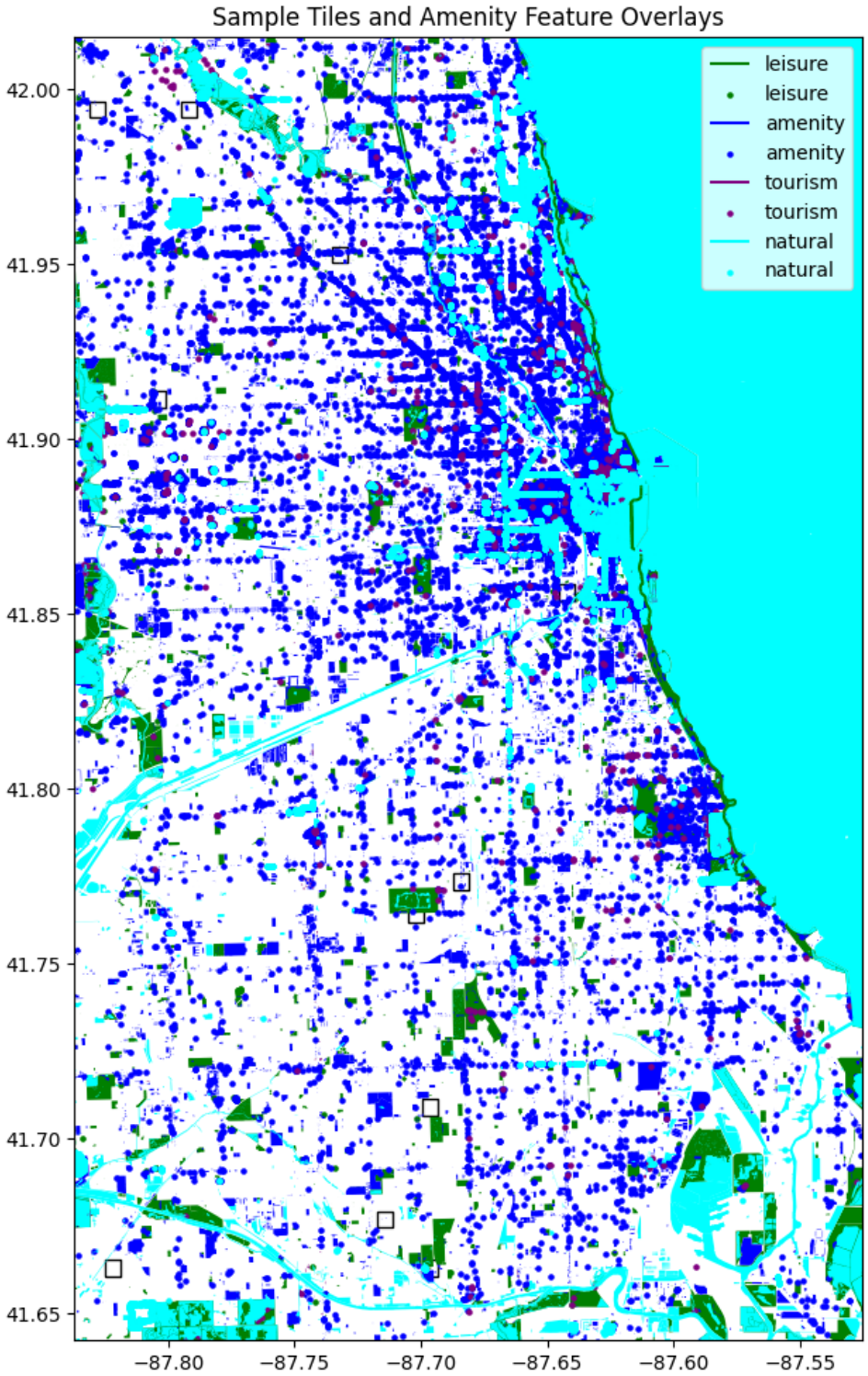


**Figure 4.** ***Sample overlay showing OSM features corresponding to tiles such as parks, schools, museums, and bodies of water. This illustrates how labels are assigned via spatial intersection.***