Urban Heat Island Detection from Satellite Imagery using Convolutional Neural Networks

DSAN5550 Project proposal: Refined version

Yuang Li

Problem definition:

Urban heat islands (UHIs) are areas in urban environments that have significantly higher temperatures than surrounding rural regions. This is primarily due to human activities, reduced vegetation, and impervious surfaces. As climate change becomes even worse, UHIs results in increased energy consumption, public health risks, and urban sustainability challenges. Detecting and analyzing UHIs is important for urban climate adaptation and planning.

This project aims to automatically identify heat-prone regions in urban areas using satellite-derived land surface temperature (LST) data and convolutional neural networks (CNNs). By detecting high-temperature regions from remote sensing imagery, we can support mitigation strategies such as targeted green space expansion or heat-resilient infrastructure planning.

Data Collection & Refinement

We will use the publicly available dataset: "A global seamless 1 km resolution daily land surface temperature dataset (2003–2020)" published by Iowa State University on Figshare. Each .tif file contains temperature values in Kelvin, where each pixel corresponds to a 1 km² area on the Earth's surface and represents the land surface temperature for that specific day.

(https://iastate.figshare.com/articles/dataset/A_seamless_1_km_resolution_global_daytime_1_30_PM_land_surface_temperature_dataset_in_2020/14885825)

In this project, we will extract and process a subset of the data—focusing on several days in 2020 and one or two selected urban regions (e.g., Washington D.C., DMV area). Each .tif file will be clipped to cover the region of interest (ROI), and then converted into smaller 2D image patches (e.g., 128×128 pixels). These patches will be treated as input images for a supervised learning task.

To create ground-truth labels for training, we will apply a temperature threshold (e.g., > 305K ≈ 32°C) to each patch. Patches where the majority of pixels exceed this threshold will be labeled as heat island (1), while the rest will be labeled as non-heat island (0). This allows us to transform the heat island identification task into a binary image classification problem.

Data refinement steps include:

• Clipping region of interest (ROI)

• Rescaling temperature values (normalization)

• Splitting into train/test/validation sets

• Generating heat island label masks

Implementation

We will use a convolutional neural network (CNN) to classify each patch as either a heat island or non-heat island, based on spatial temperature patterns. The model will be implemented in PyTorch and trained on labeled LST image patches, where labels are generated by applying a temperature threshold (e.g., 305K ≈ 32°C) to create binary supervision targets.

While the labels are derived from a simple threshold, the CNN itself is not threshold-based; rather, it learns to recognize structured spatial features—such as concentrated high-temperature clusters—that are indicative of urban heat islands. This enables the model to generalize to new data and distinguish meaningful heat island regions from isolated hot pixels or noise.

Our main focus is on patch-level classification. We may explore semantic segmentation models (e.g., U-Net) as an extension if possible, to produce pixel-wise predictions. However, we prioritize learning robust, high-level features for binary classification to identify heat island regions at the patch level.

Evaluation

Model performance will be evaluated using appropriate metrics such as accuracy, F1-score, precision, recall. At the end, the model should be able to accurately identify regions with high surface temperatures and demonstrate reasonable predictive accuracy. Also, we will visualize the detected high-temperature areas through a heatmap or a prediction overlay to clearly illustrate the results. We will use the CodeCarbon library to track and report the carbon emissions generated during model training and inference, and include the results in both the final paper and poster.