Estimating Building Energy Efficiency From Street View Imagery, Aerial Imagery, and Land Surface Temperature Data

Kevin Mayer & Lukas Haas - {kdmayer, lukhaas}@stanford.edu

Stanford | ENGINEERING
Computer Science

CS 231N: Deep Learning for Computer Vision – Spring 2022

Background & Introduction

- The **building sector** accounts for approximately **40% of the EU's energy consumption** and for **36% of its greenhouse gas emissions**.
- As of today, about **75% of the EU building stock remains energy inefficient** and **less than 1% of the national building stocks are renovated each year.**
- In the short term, the EU needs **to double the renovation rates** in order to **meet its own climate targets**

Problem Statement

- Improving the energy efficiency of existing buildings is challenging because **energy efficiency is determined via on-site visits** which are slow, costly, and geographically incomplete. **Scalable solutions are needed.**
- We propose a **novel method** which can **classify buildings into two classes namely energy efficient and inefficient ones from remotely sensed data only.**

Dataset

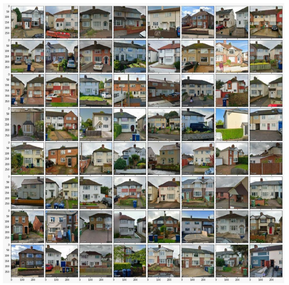
- Our dataset consists of 39,605 buildings, each of which is represented by a single **street view** and **aerial image,** an OpenStreetMap (OSM)-derived **footprint polygon**, and satellite-borne **land surface temperature** (LST) measurements.
- We collected building-level observations from **four cities across the United Kingdom:** 21,607 buildings from the city of **Coventry**, 6,834 buildings from **Westminster** in the city of London, 7,464 buildings from **Oxford**, and 3,700 buildings from **Peterborough**.
- We aggregate the building labels in a **binary** fashion, i.e. **energy efficient** (EU label "A" to "D") and **inefficient** (EU label "E" to "G").
- Based on the binary grouping, the dataset exhibits a strong **class imbalance** with **65.94%** of the buildings falling into the energy efficient group and **34.06%** into the inefficient one.

(a) Aerial Image

(b) Street View Image

Methodology

- Data Cleaning with **K-Means Clustering and Semantic Search**



(a) Clean Street Views

(b) Noisy Street Views

- Data Cleaning with **Semantic Segmentation**

(a) Original Street View

(b) Segmented Street View

- Baselines:
  - This study uses both **k-NN** and **SVM models as baselines** due to their ability to **perform well on high-dimensional data.**
  - As a third baseline, we introduce a **majority model** which simply predicts the **majority class** for each data point.
- End-to-End Deep Learning:
  - Two **Inception-v3-based image encoders**
  - **Two-layer feedforward network** trained on **joint embedding space**

Baseline Experiments & Analysis

- The **SVM model performs superior to the other two baseline methods** while the k-NN model does not generalizing well across cities different cities.
- The **end-to-end deep learning model performs best**, fusing the different data sources to achieve an **F1-score of 62.06%**
- For evaluation, **macro-averaged metrics** are chosen **to give both classes equal weight**

Table 1: Quantitative Model Results on Peterborough Test Set (in %)

Model	Binary Classification (Macro)		
	Precision	Recall	F1
Majority Model	40.55	50.00	44.78
k-Nearest Neighbor (k-NN)	51.97	51.14	50.59
Support-Vector Machine (SVM)	56.59	56.31	56.44
End-to-End Deep Learning	**61.99**	**62.86**	**62.06**

DL Experiments & Analysis

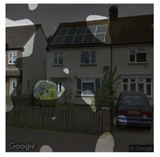
Qualitative Results - Integrated Gradients

- To augment our analysis with qualitative results, we perform the **integrated gradients attribution method** to **evaluate the joint contributions of street view and aerial images** with respect to our model's predictions.
- This study **adapts code** from the attribution framework **TruLens** to work in a **multi-data source setting**.
- Street view and aerial contributions are calculated w.r.t. random image baseline (x') using the **formula**:

$$\text{Attr}(x_i) = (x_i - x') \cdot \frac{1}{m} \sum_{k=1}^{m} \frac{\partial F(x' + \frac{k}{m} \cdot (x_i - x'))}{\partial x_i}$$

- Our work finds that while the end-to-end model generally focuses on relevant part of the images, this is not always the case.

(a) Original Street View



(b) Street View Attribution Map

(c) Original Aerial Image

(d) Aerial Attribution Map

Ablation Study examining the Predictive Power of each Data Source

- Our study finds that **all data sources contain complementary signals** correlated with building energy efficiency and **averaging the predictions of single-source models results in the best performance**.
- The results suggest that with more data and sophisticated modeling techniques, **end-to-end model performance could be further improved.**

Table 2: Ablation Study on Peterborough Test Set (in %)

Model	Binary Classification (Macro)		
	Precision	Recall	F1
End-to-End Deep Learning (2 Inception-v3 Models + 2 Linear Layers)	61.99	62.86	62.06
Inception-v3 (Street View) + Log. Regression	60.31	60.73	60.51
Inception-v3 (Aerial Imagery) + Log. Regression	59.07	58.99	59.03
Land-Surface Temperature + Building Footprint + Log. Regression	57.56	55.82	56.24
Averaging Model Predictions	**65.86**	**62.51**	**63.67**
Land-Surface Temperature + Log. Regression	57.93	55.71	56.15
Building Footprint + Log. Regression	57.71	56.01	56.45

Conclusions & Future Work

- This work presents a **novel method to to estimate binary building efficiency using remotely sensed data** sources only.
- We make **three major contributions**:
  - **First work utilizing satellite-borne heat loss data** to estimate building-level attributes such as energy efficiency
  - An **original multi-source dataset** for estimating energy efficiency
  - An **ablation study** of the predictive power of each data source
- **Future work** could focus on **improving the end-to-end deep learning model** and **explore the potential of LST data** for related tasks such as the estimation of household-level energy consumption.