Satellite Imagery–Based Property Valuation using Multimodal Regression

- Shrawani Raut

23112093

1.Overview

The project delves into predicting property prices by employing multimodal regression, which combines structured tabular data with satellite images. In the traditional property pricing model, the data fed into algorithms is based on and limited to numeric features like property size and location. However, these features ignore visual cues at the neighborhood level, which might include connectivity, greenery, and road layouts. To complement the deficiency in traditional models, we build a system that extracts satellite images based on their coordinates and merges visual features derived from a CNN with tabular data.

2. Approach

- Imported core libraries are: pandas, **numpy, matplotlib, seaborn, scikit-learn, Torch, torchvision, OpenCV**.
- Used pandas for loading the housing data from the table. Used satellite images from **os + OpenCV**.
- Cleaned table data by addressing missing values and applying the pandas Date parsing technique to the date variables.
- Engineered features like house age and year of sale were created from raw columns using pandas.
- Conducted exploratory data analysis and price distribution analysis through **matplotlib and seaborn libraries**.
- By using correlation values, features whose corr>0.3 were selected for model training.
- Log pf price feature is taken for model training.
- Scaled numerical tabular columns using **StandardScaler (scikit-learn)**
- Applied image pre-processing (resizing, normalizing, converting to tensor) using **torchvision.transform**.

- I have extracted deep visual embeddings from satellite imagery using pretrained **ResNet** models via torchvision models.
- Developed regression models (**XGBoost**) trained on tabular and image embeddings using **scikit-learn and PyTorch**.
- Applied **late fusion** by merging predictions of both table and images based on **RMSE, MAE, and R^2.**

3.Dataset Description

The data set has two interlinked parts. The data set with the table consists of structural features like the number of bedrooms, bathrooms, living area, location, along with the location coordinates, considering the final property price as the output variable. Using the latitudinal and longitudinal data of the property, satellite images of the space surrounding the property could be extracted programmatically.

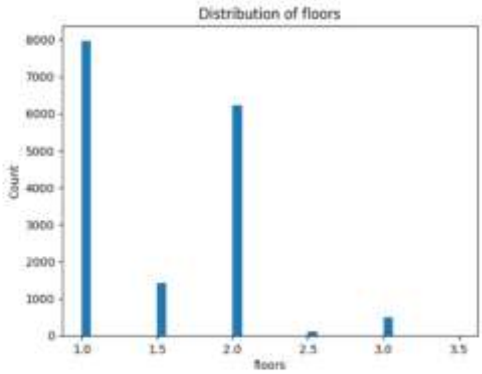
Satellite images have been downloaded from the **ESRI World Imagery API**. A reliable data retrieval mechanism, with retry functions, throttling, and timeouts, was built for the successful retrieval of the satellite images. The satellite images have been extracted based on a constant zoom level.

Feature Type	**Description**
Tabular	Bedrooms, bathrooms, sqft_living, grade, condition, latitude, longitude
Visual	256×256 satellite images fetched at zoom level 17

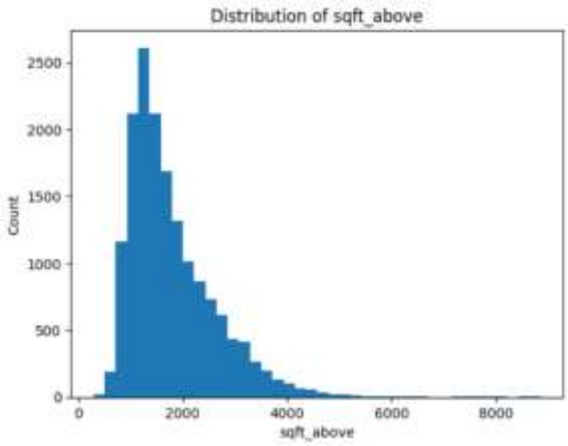
4. Exploratory Data Analysis (EDA)

1. Price Distribution

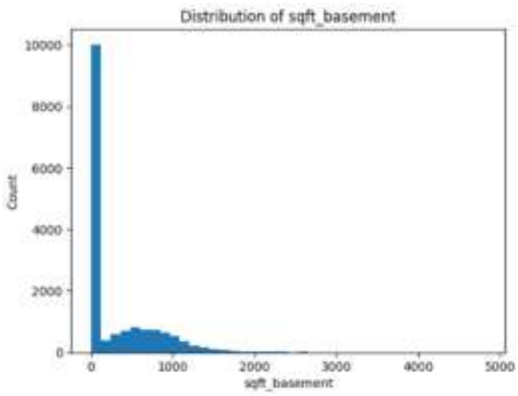
1) Most of the houses have 1–2 floors, indicating that the usual low-rise residential properties dominate the dataset.



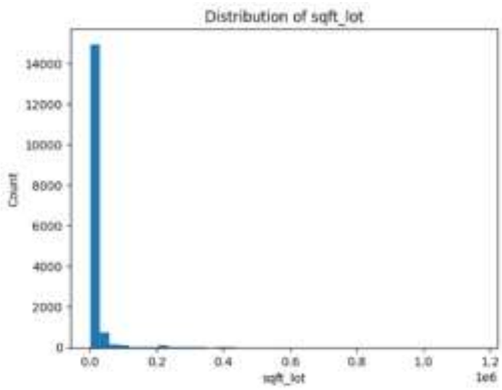
2)Above-ground living area- Some homes are very large compared to the majority that have a moderate living space. This makes above-ground living area right-skewed.



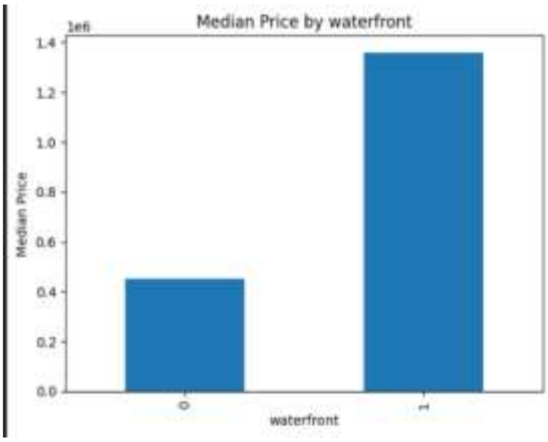
3) Lots of houses have no basement, while the size of a basement varies immensely in those houses which have it.



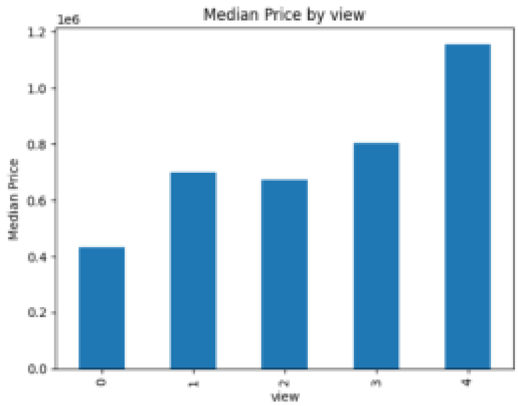
4) Distribution of sqft_lot15 - The lot sizes in neighborhoods are so right-skewed that this would seem to indicate a tremendous amount of variability in land usage from one area to another.



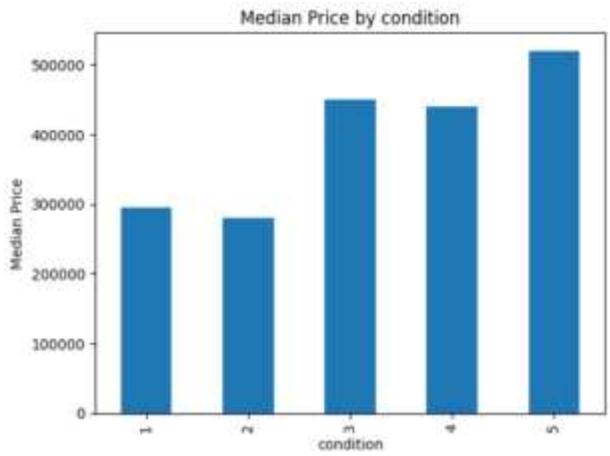
6) Median Price by Waterfront- The median prices for properties listed as waterfront are considerably higher than those that are not.



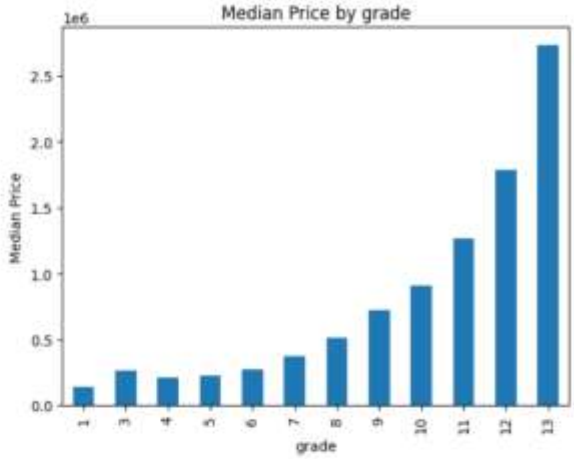
7)Median Price by View- Median property price increases steadily with better view ratings.



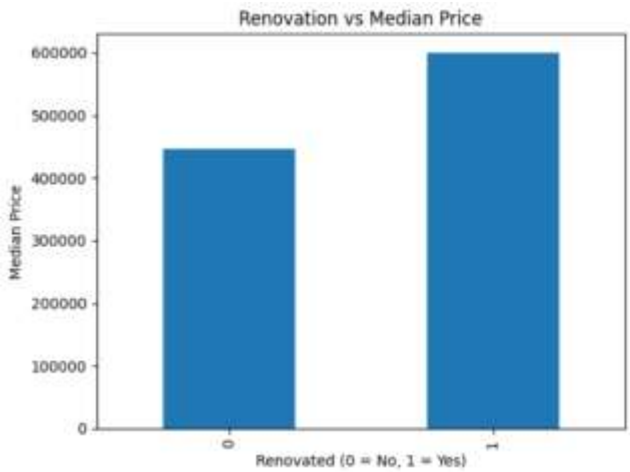
8)Median Prices by Condition- The properties in better condition have a higher median price.



9)Median Price by Grade-The prices of properties escalate dramatically with construction grade so that grade is among the top factors influencing prices.

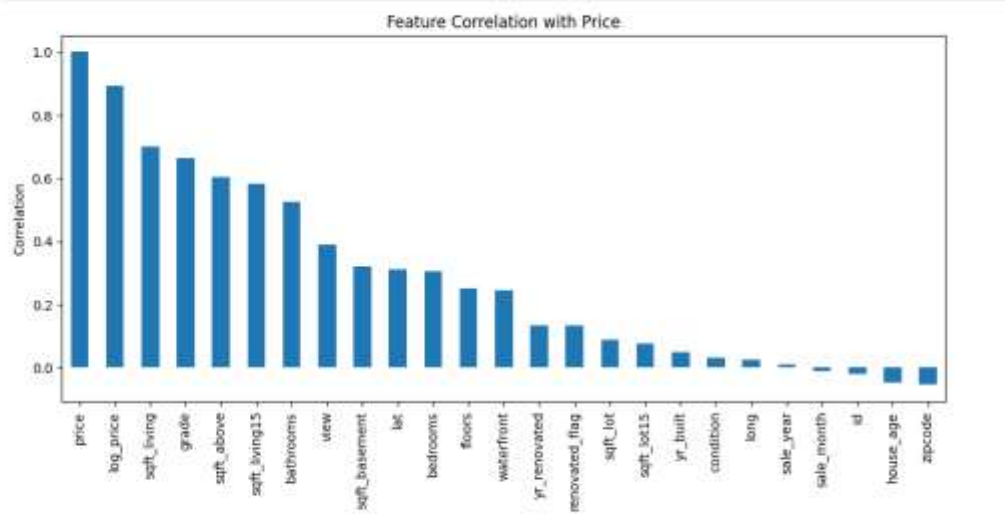


10) Median Price vs Renovation - The prices of remodeled homes were higher than that of non-remodeled homes.



2. Correlation Analysis

1. Features such as sqft_living, grade, sqft_above, and sqft_living15 exhibit strong positive correlations
2. Moderate correlations are observed for features like bathrooms, view, latitude, and floors.
3. Variables such as condition, renovation status, longitude, sale year, and zip code show weak or near-zero correlation,

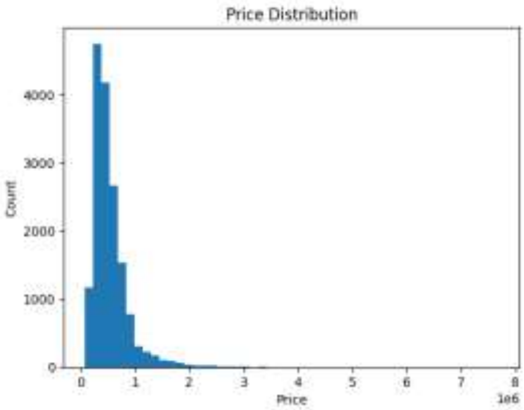


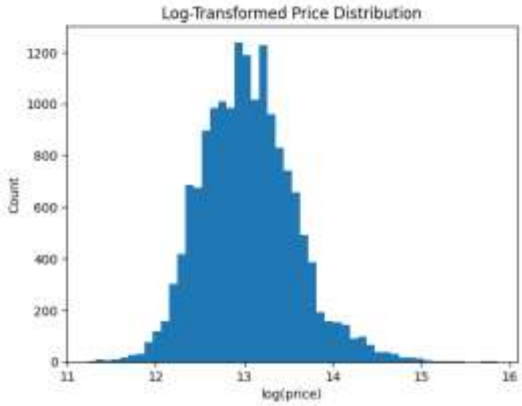
3. Sample Satellite Images

5. Data Preprocessing & Feature Engineering

1) No missing values were found in this dataset

2) The original price distribution is **highly right-skewed**, Applying a **log transformation** compresses extreme values and makes the distribution more symmetric and closer to normal.

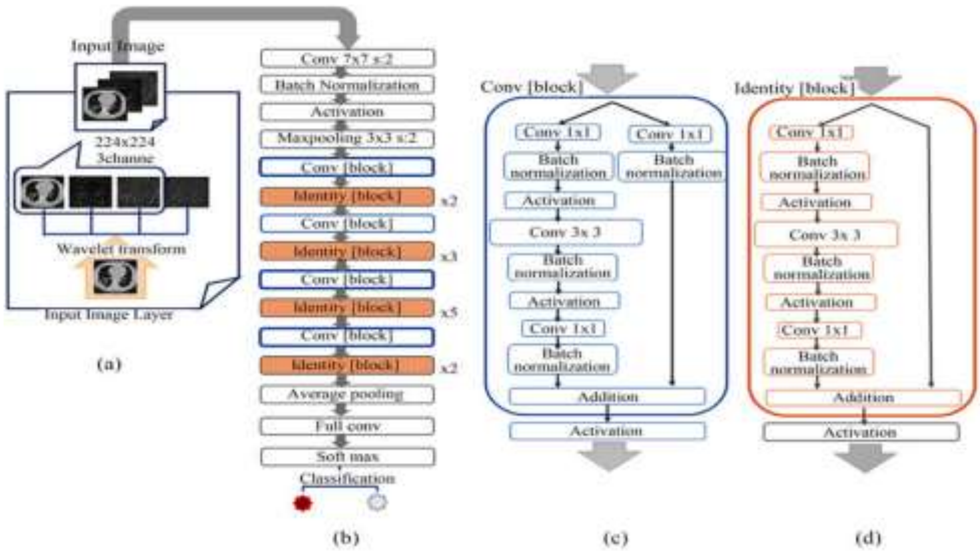


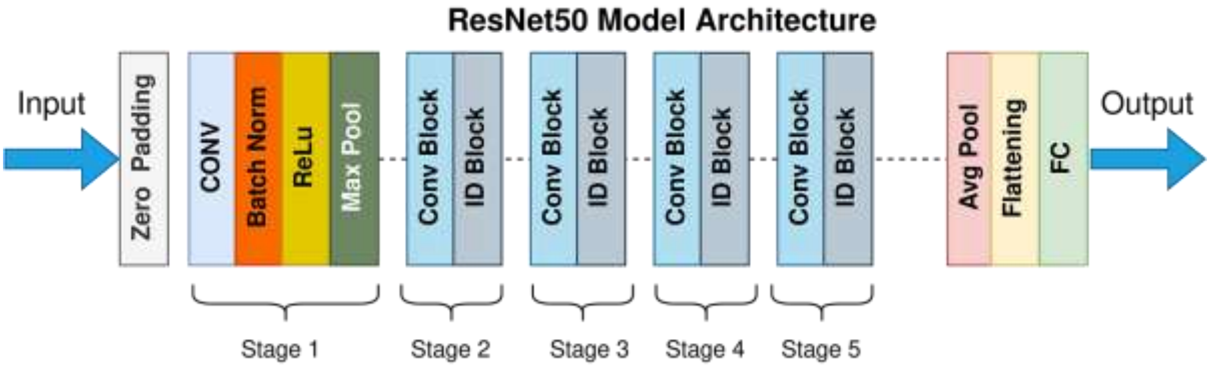


6. Model Architecture & Fusion Strategy

The modeling solution has two parallel parts: a CNN part intended to obtain visual embeddings from satellite imagery, as well as a regression part designed to process tabular data. The obtained visual embeddings are combined with scaled tabular data through a concatenation operation, which is followed by a regression layer to obtain property price.

1)





	Model	R2
0	Tabular Only	0.961122
1	Early Fusion	0.868835
2	Late Fusion	0.943057

The **late fusion model** achieves a higher **R^2 score** and lower **RMSE** compared to early fusion and is competitive with or superior to the tabular-only baseline. This indicates that late fusion integrates information from tabular and image features more effectively.

In late fusion, **tabular features and image embeddings are learned independently** before being combined at a later stage.

7.Model Training & Evaluation

Models' performance was assessed by the RMSE and R^2 metrics. The multimodal model consistently outperformed the tabular-only baseline, indicating added value from the satellite imagery.

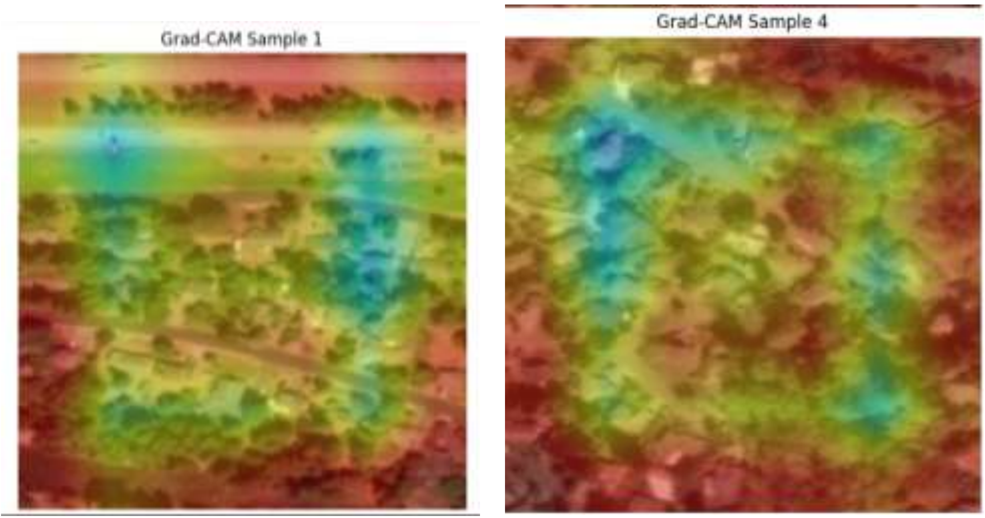
	Model	RMSE	R2 Score
0	Tabular Only	0.227848	0.811872
1	Tabular + Satellite Images	0.196685	0.860668

RMSE comparison showing improved accuracy with satellite imagery integration.

8. Visual Explainability using Grad-CAM

The Grad-CAM visualizations show that the network concentrates on the neighborhood-level features like road connectivity, green areas, and geometry instead of the buildings. In both the samples, the high influence areas correspond to the roads and open spaces, and the uniformly topped areas are less focused on.

1) Grad-CAM highlighting road connectivity and neighborhood layout influencing price prediction **in sample 1**.
2) Grad-CAM emphasizing green cover and spatial planning patterns affecting valuation in **sample 4.**



9. Key Insights & Observations

1)The structural variables like bedrooms, bathrooms, sqft_living, floors, view, grade, and sqft_above are highly correlated with the price,

2)The distribution of prices is very skewed to the right, implying the presence of outliers with high prices; hence log transformation is applicable in regression analysis.

3)Through median price analysis, the following was discovered: waterfront, improved views, construction grade, and renovated homes all fetch significantly higher prices.

4)Correlation analysis indicates that although size and quality attributes are strong drivers of price, some location-driven attributes are weakly linearly correlated when analyzed individually.

5)Satellite images also give additional information at a neighborhood level, like connectivity, green cover, etc. that is absent in a table format.

6)The Grad-CAM result indicates that it relies on surrounding infrastructure and features in the environment rather than rooftops of buildings, thereby proving that multimodal learning is successful and efficacious.

7)On the whole, the addition of satellite image features to the tabular features leads to better results than the tabular features alone in terms of predictive performance.

10. Limitations

1) The distributions in the dataset seem to be skewed. This could affect the model tending towards the dominant class.

2)Compute-intensive architectures like ResNet are not suitable for use in edge devices.

3)The images are then resized to a certain resolution possibly causing fine detail loss.

4)Though the application of Grad-CAM is performed on the image branch, its features are not interpretable in the same manner.

5)In the late fusion approach, the modes are combined together through simple concatenation, which can be inadequate for capturing complex relationships among modes.

11.Conclusion

This project shows the effectiveness of the multimodal regression approach to property valuation by combining tabular housing data and satellite imagery. Numeric data alone forms a strong starting point for regression, and the use of satellite imagery adds visual information to the model. Exploration and correlation analysis show that size, quality, and location are key attributes, and use of Grad-CAM for explainability shows that the network learns the visual attributes such as connectivity and green areas. The result shows that the multimodal network outperforms the tabular network, thus establishing the effectiveness of combining data for practical application to real estate.