Satellite Imagery–Based Property Valuation

1. Overview: Approach & Modeling Strategy:

Traditional house price prediction models rely heavily on structured tabular attributes such as square footage, number of bedrooms, or location coordinates. However, these features often fail to capture environmental and neighborhood context, such as greenery, road density, water proximity, or urban congestion.

This project proposes a multimodal regression framework that integrates:

- **Tabular Data**: Property attributes (e.g., sqft_living, bedrooms) and engineered spatial features such as dist_from_center.
- **Satellite Imagery**: High-resolution tiles captured via the **Mapbox Static API** to provide visual and environmental context.

Key Idea

The core innovation is learning complementary representations from structured data and satellite images, then fusing them adaptively using a **learnable gating mechanism**. This allows the model to dynamically weight the importance of the visual context against the physical property specs for each individual valuation.

High-Level Pipeline

1. **Automated Data Acquisition**: Programmatically retrieve *224 x 224* satellite images centered on property coordinates using data_fetcher.py.
2. **Feature Engineering**: Process raw data through preprocessing.py to calculate house_age, relative_size, and generate a spatial center for distance-based features.
3. **Visual Embedding Extraction**: Utilize a pretrained **EfficientNetB0** backbone to transform raw pixels into high-dimensional spatial embeddings.
4. **Gated Multimodal Fusion**: Integrate both streams using a sigmoid-activated gate that computes a weighted sum of tabular and visual signals.
5. **Robust Regression**: Predict log-transformed prices (*price_log*) using **Huber Loss** to minimize the impact of outliers during training.

2. Grad-CAM (Visuals and Financial Insights):

2.1 Methodology:

For each property:

- The trained multimodal model (*EfficientNetB0 + tabular features*) was used.
- Real tabular data (bedrooms, sqft, grade, location, etc.) was provided.
- Gradients were computed from the last convolutional layer (*top_conv*).
- The resulting heatmap was overlaid on the original satellite image. This ensures that the explanation corresponds to the actual prediction made by the model, not a dummy or isolated image-only prediction.

2.2 Interpretation of heatmap colors:

Color	Meaning
🔴 Red / 🟡 Yellow	Strong influence → Higher predicted price
🟢 Green	Moderate contribution
🔵 Blue	Low or negligible contribution

Important: Colors show *relative importance*, not absolute monetary value.

2.3 Key observations across all images:

2.3.1 Neighborhood Matters More Than Just the House:

The model consistently focuses on:

- Road connectivity & intersections
- Open land and plot size
- Green cover & tree density
- Spacing between nearby properties

➔ Indicates learning of neighborhood-level valuation signals, crucial in real estate pricing.

2.3.2 Environment Drives Value Signals:

Instead of only rooftops, the CNN captures:

- Accessibility
- Openness
- Surrounding infrastructure

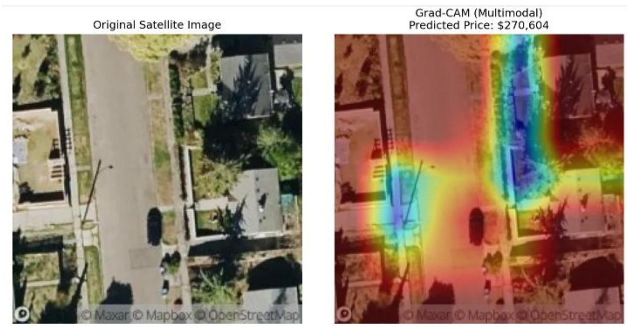
➔ Exactly the kind of contextual information satellite imagery is expected to provide.

2.3.3 Prediction Errors Are Explainable:

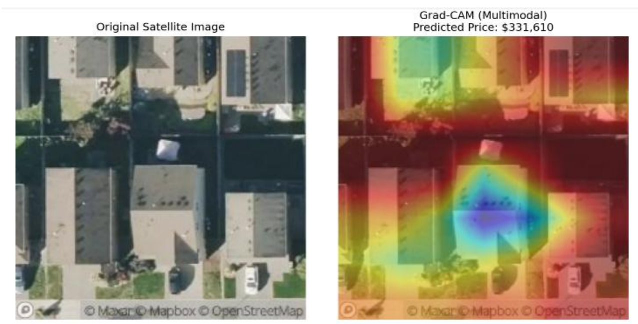
Differences between predicted and actual prices may arise due to:

- Interior quality & renovations (not visible)
- Market negotiation effects
- Temporal, legal, or policy factors

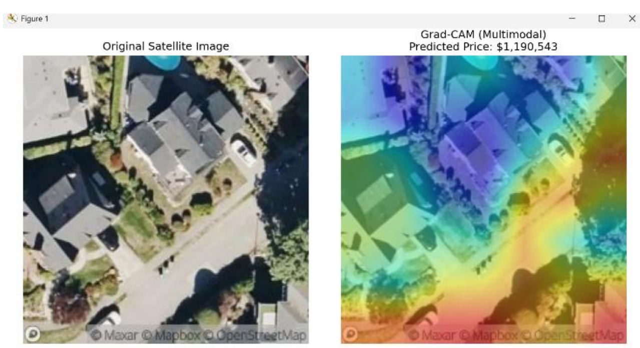
➔ Despite this, Grad-CAM remains consistent and interpretable, indicating stable learned patterns rather than noise.



Actual price->220000
Left: Actual satellite image
Right: gradCAM generated image
Figure 1(train_79): Near-accurate prediction case. The model slightly over-predicts price by emphasizing road connectivity and open surroundings. Grad-CAM highlights meaningful neighborhood features, indicating that prediction differences arise from non-visual factors rather than model failure.

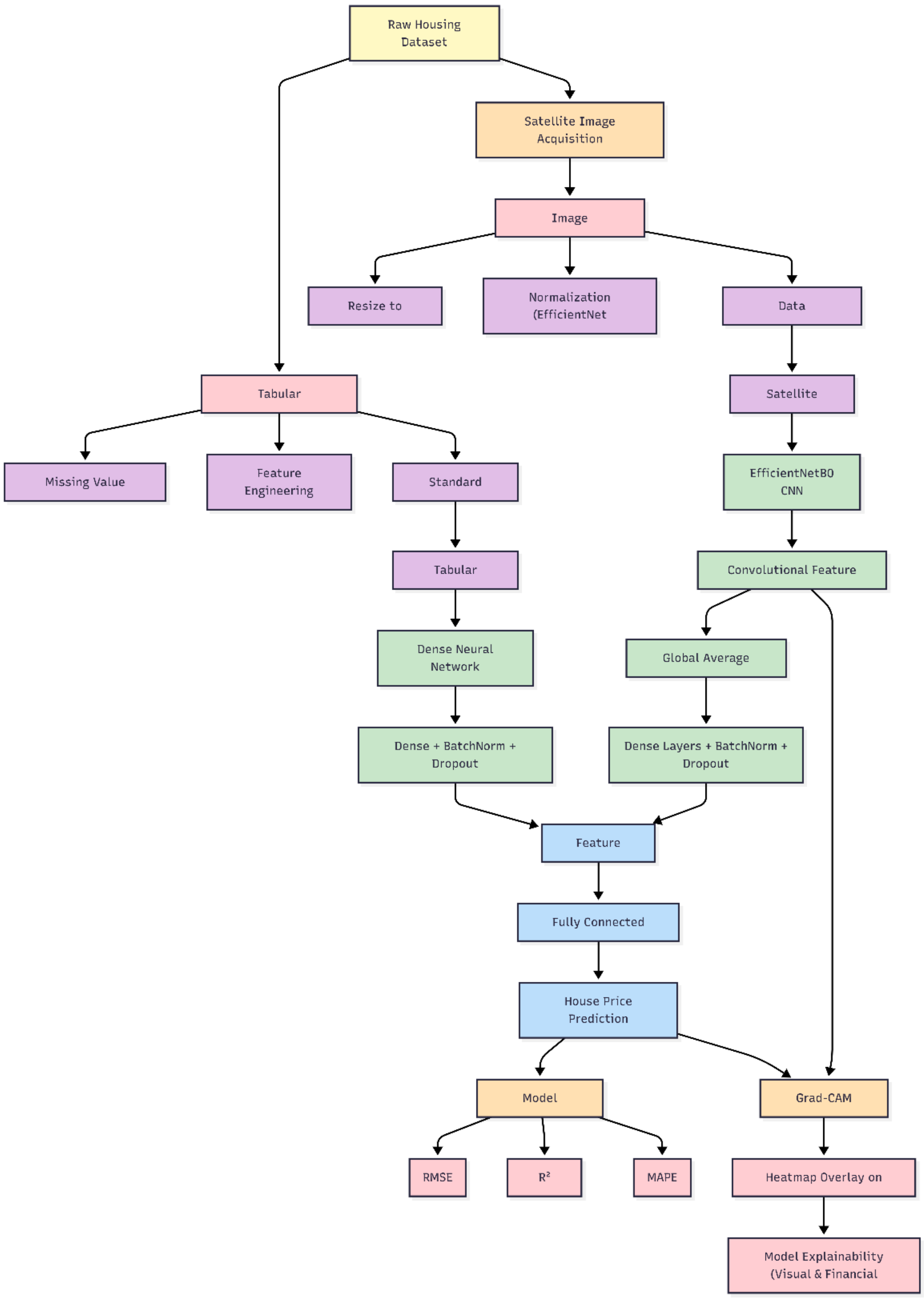


Actual price->320000
Left: Actual satellite image
Right: gradCAM generated image
Figure 2(train_242): Dense residential property case. The heatmap highlights clusters of rooftops and surrounding built-up density, indicating that the model relies on neighborhood-level features such as housing density and urban layout, rather than a single house, to inform its price prediction



Actual price-> 1400000
Left: Actual satellite image
Right: gradCAM generated image
Figure 3(train_308): High-value property in a dense residential neighborhood case. The heatmap concentrates on the house footprint, surrounding greenery, and road connectivity, indicating that the model associates well-developed residential layout, landscaping, and neighborhood quality with higher property prices.

3. Architecture Diagram:

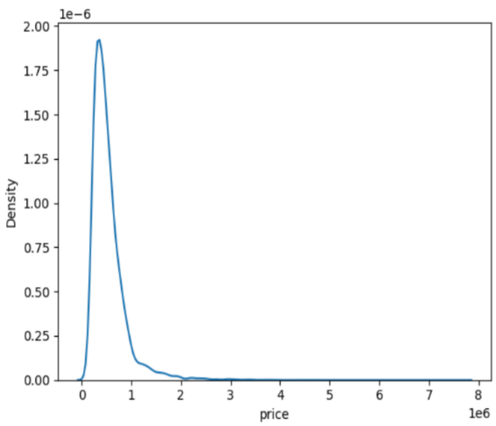


4. Exploratory Data Analysis(EDA):

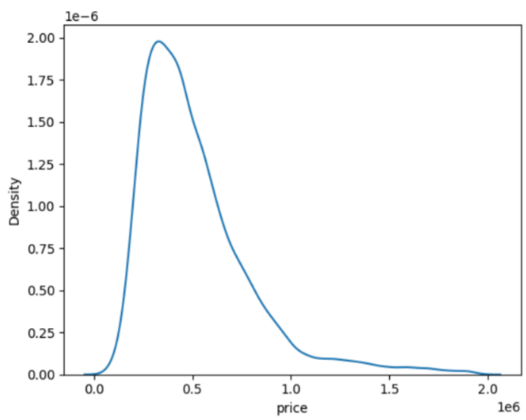
4.1 Price Distribution:

The price distribution is highly right-skewed with a long tail of high-value properties. Such skewness can bias regression models toward extreme values. Therefore, a log(1 + price) transformation was applied to stabilize variance and improve model learning.

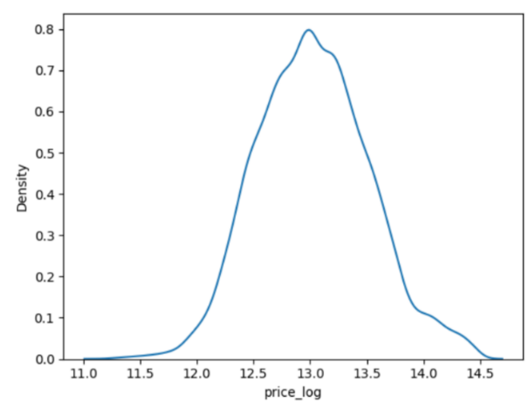
Before applying the logarithmic transformation, extreme price outliers were handled using a quantile-based approach.



Original



After removing outliers



Final(price_log)

4.2 Engineered Feature Validation:

- *house_age*: Captures depreciation effects and justifies its inclusion as a structural feature rather than relying solely on raw construction year.

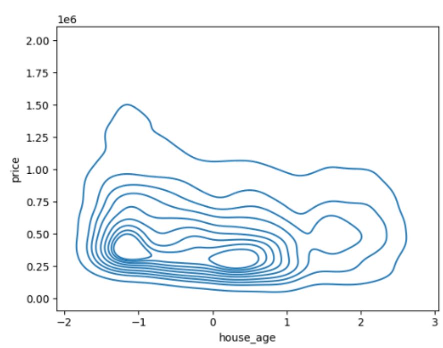


Figure: price vs house_age

The raw price distribution is highly right-skewed and dispersed, with extreme high-value outliers obscuring clear trends across house ages. This significant variability necessitates target transformation and outlier handling to ensure stable model convergence.

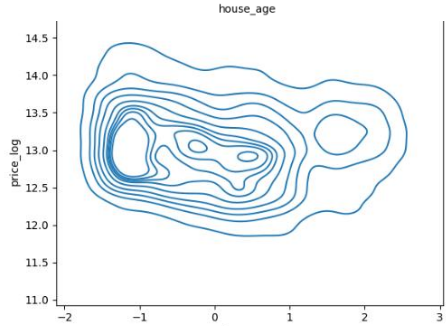


Figure: price_log vs house_age

The log transformation normalizes the price distribution, creating a stable, learnable structure for regression. The resulting relationship is more interpretable: newer houses command a premium, while older properties show a gradual, consistent decline due to depreciation.

- *relative_size*: This feature effectively normalizes property size by local context, making it more informative than absolute square footage alone.

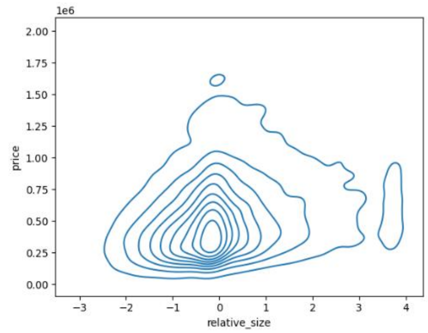


Figure: price vs relative_size

Raw prices show high dispersion and right skew across relative size values, with extreme properties dominating the upper range. This makes it difficult to observe a consistent relationship. The plot indicates the need for target transformation to stabilize variance

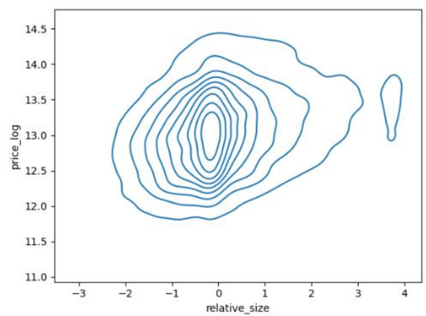


Figure: price_log vs relative_size

After log transformation, a clearer positive relationship emerges between relative size and price. Properties larger than their neighborhood average tend to have higher values. The transformed scale reveals a compact and learnable pattern suitable for modeling.

4.3 Heatmap and its interpretation:

The initial correlation heatmap reveals significant multicollinearity among raw features, particularly between size-related variables like sqft_living, sqft_above, and sqft_living15, as well as grade. This redundancy with the target variable (price) increases the risk of model instability and overfitting. Furthermore, the weak correlations exhibited by raw spatial(lat, long) and temporal (yr_built, yr_renovated) attributes suggest that these features, in their original form, fail to capture critical location or age-related trends effectively.

Figure: Before EDA

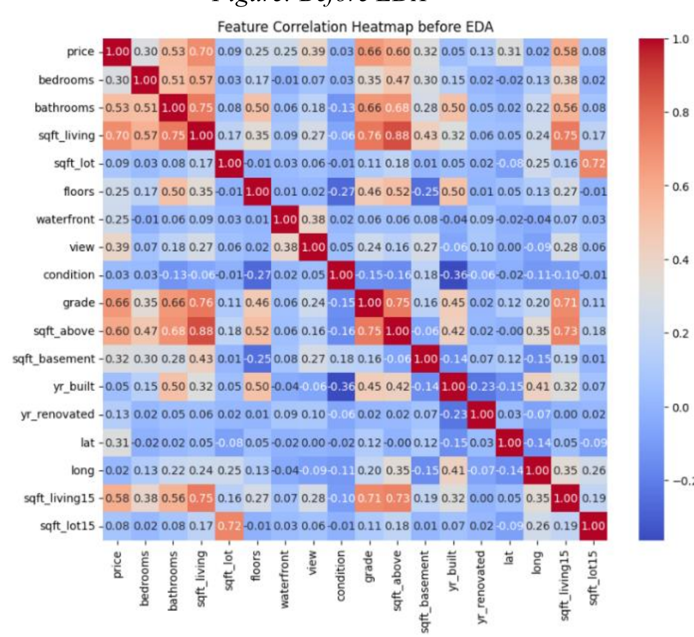
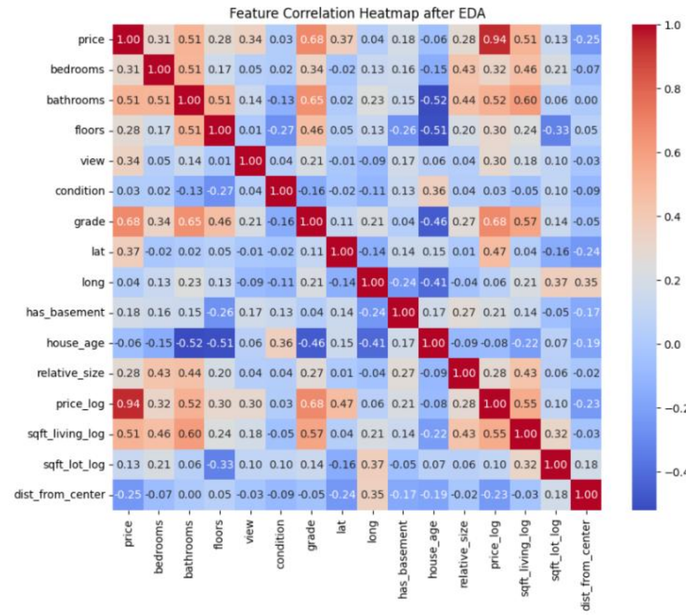


Figure: After EDA



Post-preprocessing, the correlation structure exhibits enhanced balance and interpretability, as redundant variables were systematically pruned or transformed via logarithmic scaling to mitigate multicollinearity while preserving high-variance signals. Engineered features—specifically house_age, relative_size, and dist_from_center—demonstrate distinct, moderate correlations with the target, confirming they provide complementary spatial and temporal context rather than duplicating existing data. Furthermore, the transition to price_log stabilized linear relationships with primary predictors, validating that outlier handling effectively neutralized market noise to provide a smoother optimization gradient for the Huber Loss function.

4.4 Geospatial EDA:

The geospatial distribution of property values confirms that location is the primary driver of market variance. Our analysis reveals three key findings that justify the multimodal CNN approach:

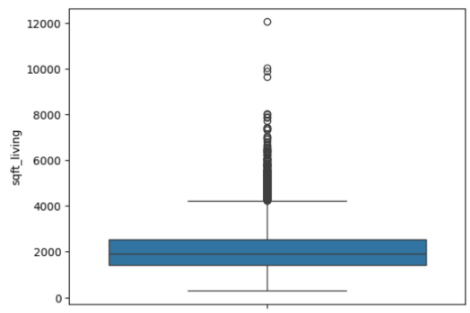
- **Spatial Autocorrelation**: High-value properties (*price_log > 13*) form contiguous clusters, showing that a home's value is deeply tethered to its neighbors—a relationship the model captures through shared geographic features.

- **Environmental Premiums**: Hotspots align with landmarks like Lake Washington and Puget Sound. The model identifies these "blue and green spaces," capturing massive visual premiums that standard tabular data cannot fully explain.
- **Contextual Segmentation**: Density contours distinguish between high-density urban zones and luxury suburbs. This allows the **Gated Fusion** mechanism to prioritize visual context in areas where "curb appeal" and aesthetics drive the valuation.

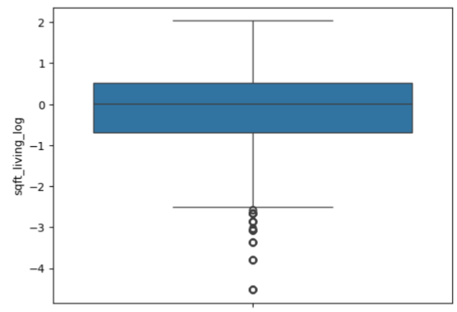
4.5 Outliers Analysis:

4.5.1 sqft_living:

Before performing EDA (left)



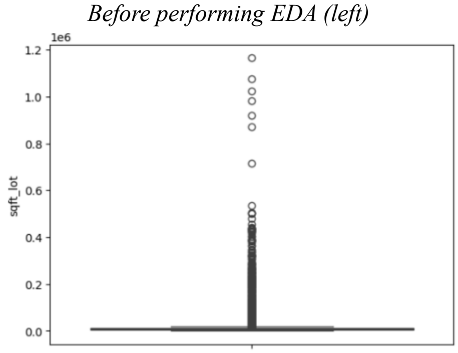
After performing EDA (right)



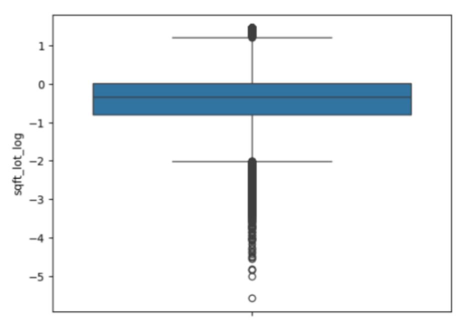
The *sqft_living* feature exhibits strong right skewness with several high-value outliers. Applying a logarithmic transformation significantly reduces skewness and compresses extreme values, resulting in a more stable distribution while retaining all valid observations.

4.5.2 sqft_lot:

Before performing EDA (left)



After performing EDA (right)



The *sqft_lot* feature exhibits extreme right skewness with substantial outliers. Log transformation effectively stabilizes the distribution and reduces scale dominance, making the feature suitable for regression modeling without discarding valid observations.

5. Results and Performance comparison:

5.1 Tabular-Only Model Performance:

Several regression models were trained using only structured tabular features, following extensive preprocessing, outlier handling, feature engineering, and scaling.

Model	R^2 (Accuracy)	RMSE (log price)	Remarks
Linear Regression	0.763	0.249	Baseline linear model
Ridge Regression	~0.763	~0.249	Similar to linear regression
Lasso Regression	~0.763	~0.249	No significant improvement
Decision Tree	~0.82	~0.21	Captures nonlinearity but prone to overfitting
Random Forest	0.880	0.180	Strong ensemble performance
Extra Trees	0.870	0.185	Comparable to Random Forest
XGBoost	0.899	0.160	Best tabular-only model

- The tabular-only models establish a strong baseline for property valuation, with ensemble methods significantly outperforming linear approaches.
- While advanced models such as Random Forest, Extra Trees, and XGBoost effectively capture nonlinear relationships among structured features, their performance plateaus due to the absence of explicit neighborhood and environmental context.
- This highlights the inherent limitation of relying solely on tabular data for real estate valuation.

5.2 Multimodal Model Performance:

The multimodal model was evaluated using the same train–validation split, preprocessing pipeline, and target transformation as the tabular-only models to ensure a fair comparison. Performance was measured using **R^2** (accuracy) and **RMSE** on the log-transformed house price.

- **R^2** (accuracy): 0.8528
- **RMSE**(log price): 0.1941

The multimodal model achieves lower accuracy and higher error compared to best performing tabular-only model. Possible reasons for this can be:

- Engineered tabular features capture most of the predictive signals in the dataset, limiting the additional benefit of satellite imagery.
- Fixed-zoom satellite images provide limited fine-grained details relevant to property value (interior quality, building condition, floors).
- Visual inputs may introduce background noise unrelated to pricing.
- Multimodal architecture increases model complexity due to increased parameters, raising the risk of overfitting.

Although quantitative performance is lower, the multimodal model provides **important qualitative advantages**:

- Enables **visual interpretability** through Grad-CAM
- Demonstrates how neighborhood context influences predictions
- Produces economically meaningful visual explanations aligned with real-estate intuition

These aspects are not available in tabular-only models.