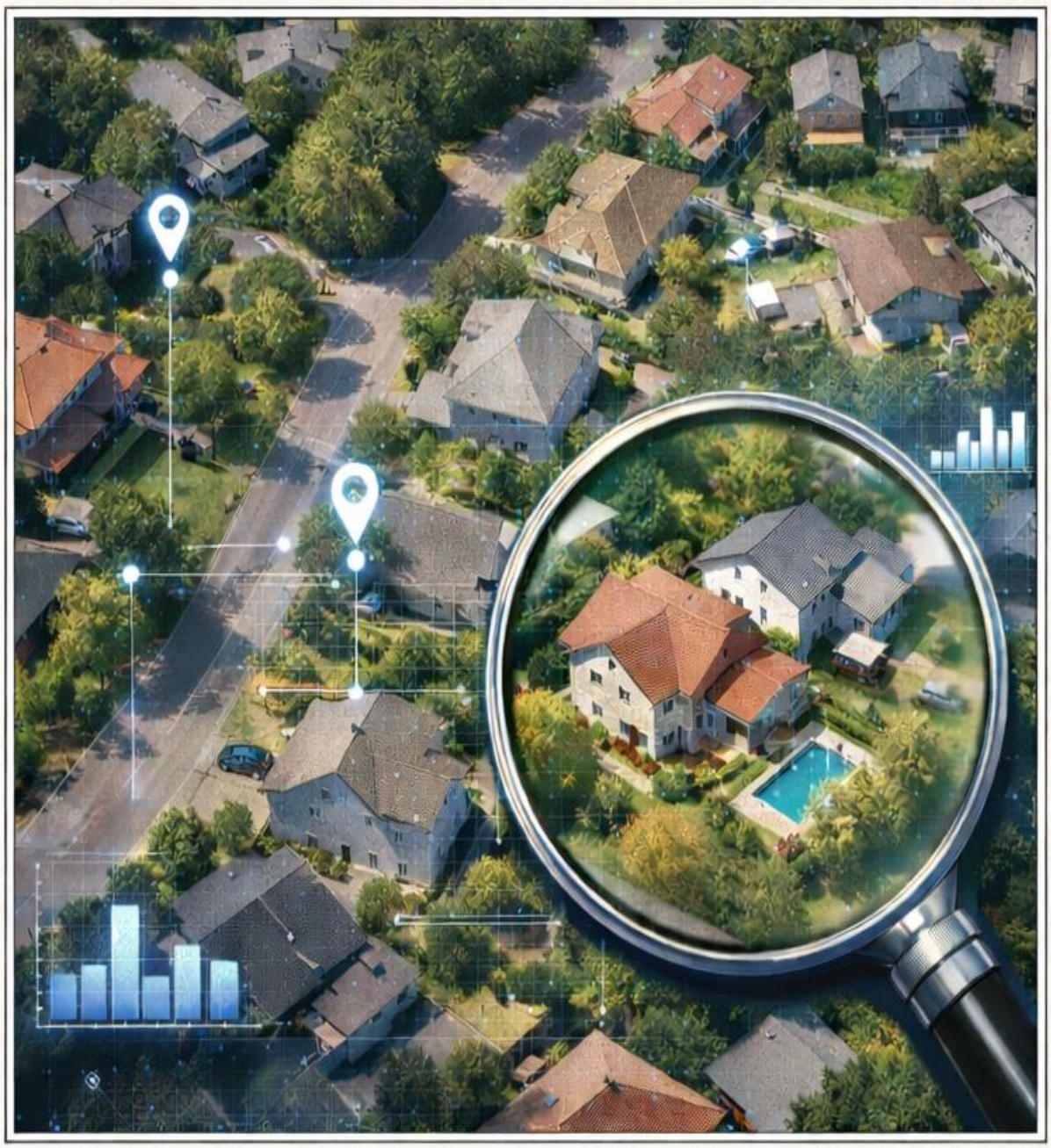
HOUSE PREDICTION MODEL

USING SATELLITE IMAGERY

PROJECT REPORT



Saral Gupta

Overview and Modelling Strategy

- This project addresses the problem of **house price prediction using satellite imagery**, formulated as a supervised regression task. The central hypothesis is that **spatial and visual characteristics visible in satellite images encode economically relevant information**—such as urban density, infrastructure development, and surrounding land use—that can complement traditional tabular data for property valuation.
- To test this hypothesis, we designed and evaluated multiple models of increasing complexity, culminating in a **hybrid deep learning architecture** that integrates visual and non-visual features.

1.1 Problem Definition and Assumptions

- The objective of this project is to predict residential house prices using **satellite imagery**, formulated as a **supervised regression problem**. Given a satellite image centered at a property location, the model learns a mapping to the corresponding house price.
- House prices exhibit a highly skewed distribution; therefore, the target variable is **log-transformed** to improve numerical stability and convergence during training.

- **The following assumptions are made:**
  - Satellite images capture spatial and structural information (e.g., building density, road networks, land cover) that correlates with property value.
  - Visual features extracted by the CNN can be effectively mapped to continuous price values through a regression head.

1.2 Tools, Libraries, and Frameworks Used

- **Python** as the primary programming language
- **TensorFlow / Keras** for deep learning model implementation
- **NumPy & Pandas** for numerical computation and data handling
- **Matplotlib & OpenCV** for visualization and heatmap overlays
- **Grad-CAM** for model explainability and spatial attribution

1.3 Data Preprocessing Strategy

Image Data

- Satellite images were resized to a fixed spatial resolution compatible with CNN input layers.
- Pixel intensities were normalized to [0,1].
- Images were loaded using a streaming pipeline to handle large dataset sizes efficiently.

Target Variable

- House prices were transformed using a **logarithmic transformation**:

$$Y = \log(\text{price})$$

- This reduces skewness and prevents dominance of high-priced outliers during training.

1.4 Model Architecture and Strategy

Baseline Model

- As a reference, a **tabular-only regression model** was trained to establish baseline performance.

CNN-based Model

- A Convolutional Neural Network was used to extract spatial features from satellite images:
- Convolutional layers learn hierarchical spatial representations
- Global pooling layers reduce dimensionality
- Dense layers map extracted features to price predictions

Hybrid Modelling Strategy

- In the final model, **CNN-derived image embeddings** were concatenated with tabular features (when applicable) and passed through a regression head. This allowed the model to jointly reason over visual and numerical information.
- The modelling strategy deliberately separates:
  - **feature learning (CNN)**
  - **price estimation (regression head)**
- to improve modularity and interpretability.

1.5 Training Procedure

- Loss function: **Mean Squared Error (MSE)** in log-price space
- Optimizer: Adam
- Training was performed in mini-batches
- Early stopping was used to prevent overfitting
- Model checkpoints were saved for reproducibility
- Performance was evaluated using:
- **RMSE**
- **R^2 score**
- computed consistently in log space.

1.6 Explainability Strategy (Grad-CAM)

- To ensure the model does not behave as a black box, **Grad-CAM** was employed on the final convolutional layer of the CNN. This method uses gradients of the predicted price with respect to convolutional feature maps to generate **spatial heatmaps**, highlighting image regions that most strongly influenced predictions.
- This step serves two purposes:

  Model validation — verifying that predictions rely on meaningful spatial structures
  Interpretability — enabling visual explanation of model behavior

1.7 Challenges Encountered and Mitigation

1.Training Instability

- Extremely negative R^2 values were observed
- Root cause: prediction collapse and target scaling issues
- Solution: strict consistency between training and evaluation spaces

2. Poor Grad-CAM Heatmaps

- Initial heatmaps were uniform or noisy
- Causes:
  - incorrect layer selection
  - vanishing gradients
- Solution:
  - Grad-CAM applied strictly to the last convolutional layer
  - regression-specific gradient handling

3. Large Dataset Handling

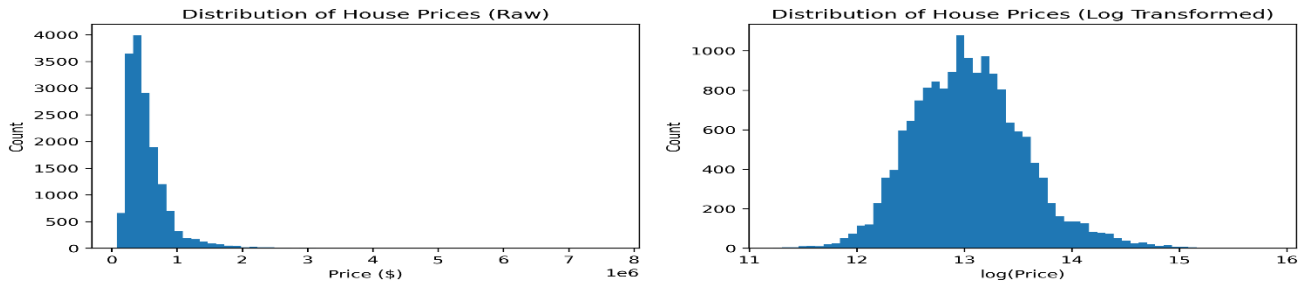
- High memory usage during image loading
- Solution:
  - batched data pipelines
  - on-the-fly preprocessing

1.8 Final Modelling Philosophy

- The final modelling strategy prioritizes:
- Learning meaningful visual representations
- Maintaining numerical stability
- Ensuring interpretability alongside accuracy
- Rather than optimizing performance alone, the focus is on building a transparent, explainable, and technically sound system suitable for real-world housing analysis.

Exploratory Data Analysis

1.Target Variable Analysis



- We began the exploratory analysis by examining the distribution of house prices, which is the target variable of our prediction task. Three complementary visualizations were used:

  1) The **raw price distribution**, showing the full range of house prices.

  2) A **log-transformed price distribution**, obtained using a log(1 + price) transformation.

  3) A **clipped price distribution**, where extreme high-price outliers were truncated to better visualize the dense mid-range of the data.

- These plots together provide a complete view of the target variable at different scales.

Why this analysis was necessary ?

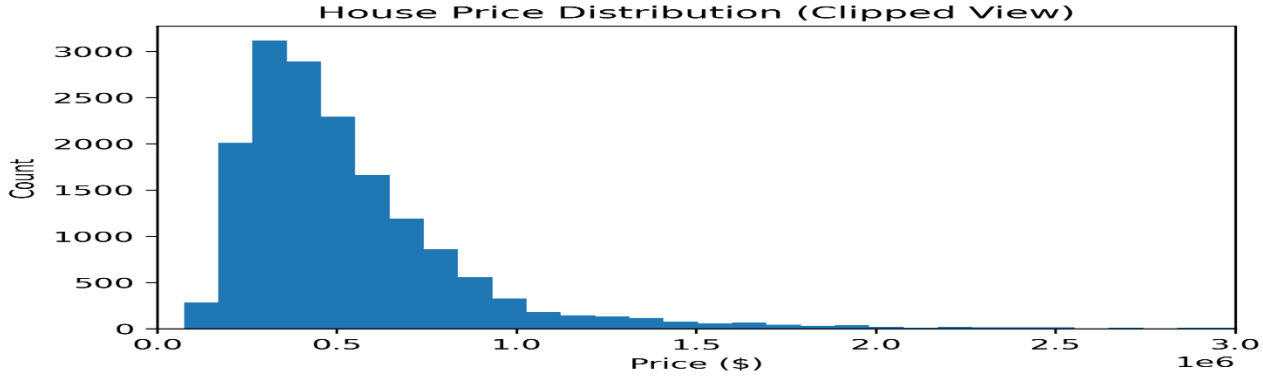
- House prices typically span several orders of magnitude, and directly modeling raw prices can lead to numerical instability and biased learning toward expensive properties. The raw distribution reveals a strong right skew with a long tail of luxury houses, indicating the presence of extreme outliers. Such skewness violates common modeling assumptions and can cause regression models to overfit high-value samples.

- The log-transformed distribution was therefore examined to assess whether the transformation normalizes the target variable. Additionally, the clipped distribution allows us to focus on the majority of homes without being visually dominated by rare, extremely expensive properties.

Key observations and modeling implications

- The raw price distribution is highly right-skewed, with most houses clustered in the lower price range and a small fraction extending into multi-million-dollar values. After applying the log transformation, the distribution becomes approximately symmetric and bell-shaped, indicating reduced skewness and stabilized variance. This confirms that predicting **log(price)** is more suitable than predicting raw price values.

- The clipped view highlights that most houses lie below a certain price threshold, reinforcing that extreme prices are rare but influential. Based on these observations, we adopt **log-transformed prices as the training target** for all subsequent models. This choice improves optimization stability, reduces the impact of outliers, and enables fairer learning across different price ranges.



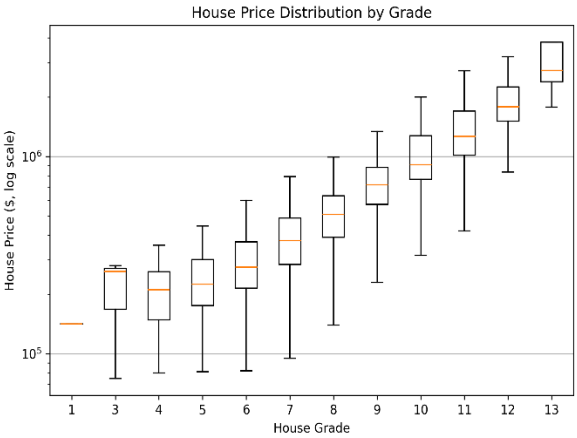
2. Relationship Between Key Features and Price

Living Area vs Price

We first analyze the relationship between living area and house price. The scatter plot (log-scaled price) shows a strong positive correlation, indicating that larger homes tend to command higher prices. However, the spread increases for larger living areas, suggesting diminishing returns—beyond a certain size, additional area contributes less consistently to price. This non-linear behavior motivates the use of non-linear models rather than simple linear regression.

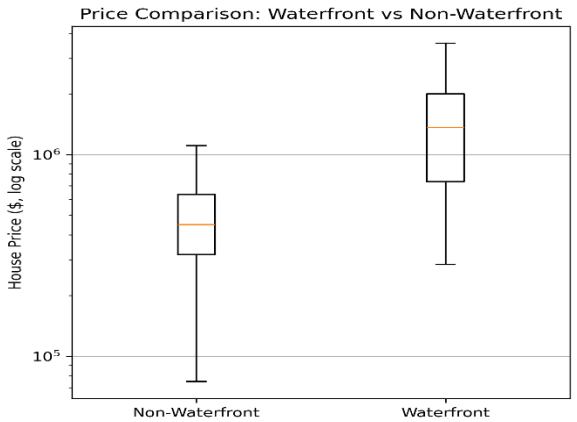
House Grade vs Price

Next, we examine house grade, which captures construction quality and design. The box plot reveals a clear, monotonic increase in median price with grade, along with tighter variance at higher grades. This indicates that grade is a highly informative categorical feature and a strong proxy for overall property quality, making it a critical input for both tabular and hybrid models.

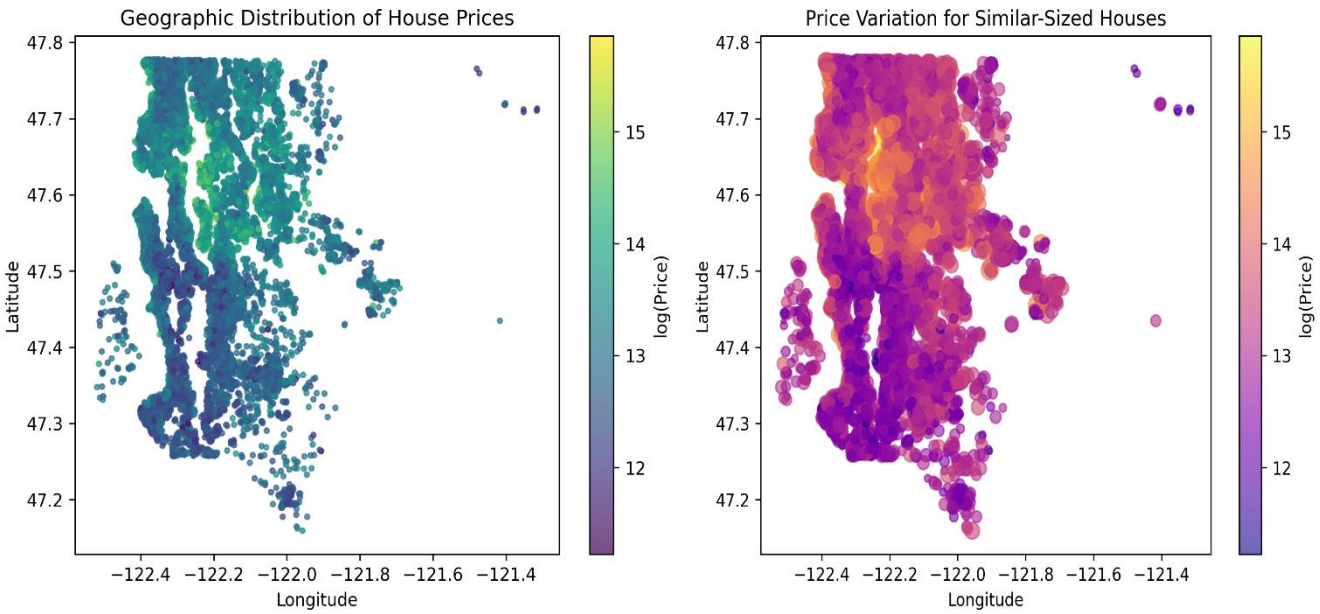


Waterfront Effect

Finally, we compare waterfront and non-waterfront properties. The price distributions show a substantial upward shift for waterfront homes, even in log space. This confirms that waterfront presence introduces a discrete price premium that cannot be explained by size or quality alone, reinforcing the importance of explicitly including this feature in the model.



3: Spatial Location Effects on House Prices



- **What we analyze.**
  We examine the geographic distribution of house prices using latitude and longitude to understand whether location introduces systematic price patterns beyond structural features such as size or grade.

- **Geographic Price Clustering.**
  The first plot visualizes house prices (log-scaled) across geographic coordinates. Clear spatial clustering is observed, where nearby properties tend to share similar price ranges. High-value regions form contiguous clusters rather than appearing randomly scattered. This indicates that location exerts a strong, neighborhood-level influence on pricing, reflecting factors such as access to amenities, waterfront proximity, and urban desirability. Importantly, this spatial structure cannot be fully captured by latitude and longitude as independent numerical features.

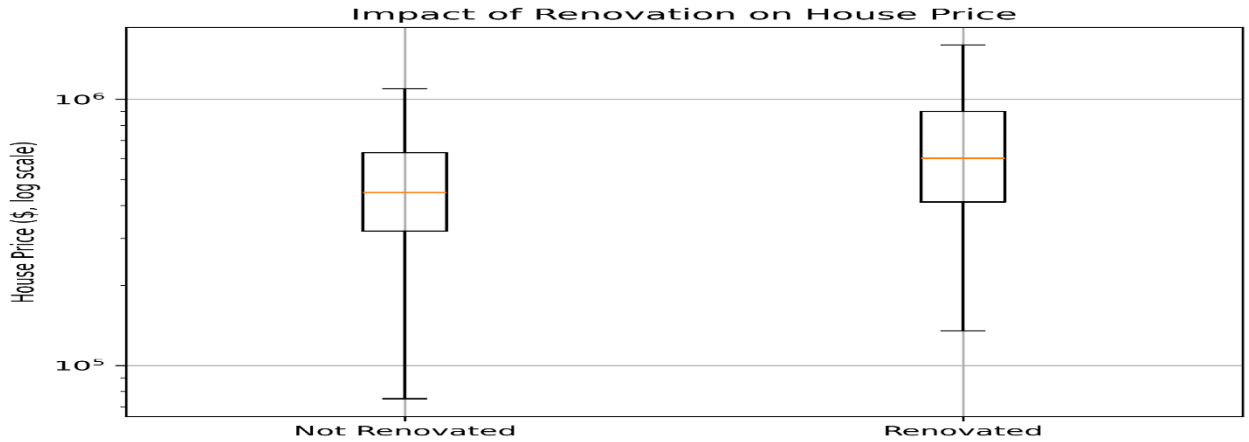
- **Price Variation Among Similar-Sized Houses.**
  The second plot focuses on houses of comparable size and shows substantial price variation across different locations. Even when controlling for living area, properties in certain regions command significantly higher prices than others. This highlights that structural attributes alone are insufficient to explain valuation differences and that surrounding context plays a major role.

- **Key Inference.**
  These observations motivate the inclusion of satellite imagery in the modeling pipeline. While tabular features encode *what the house is*, satellite images provide contextual information about *where the house is*—including neighborhood density, road layout, green cover, and proximity to water. A CNN-based image encoder is therefore well-suited to capture these spatial and visual cues that are otherwise difficult to model explicitly.

4. Effect of Renovation and Property Age on Price

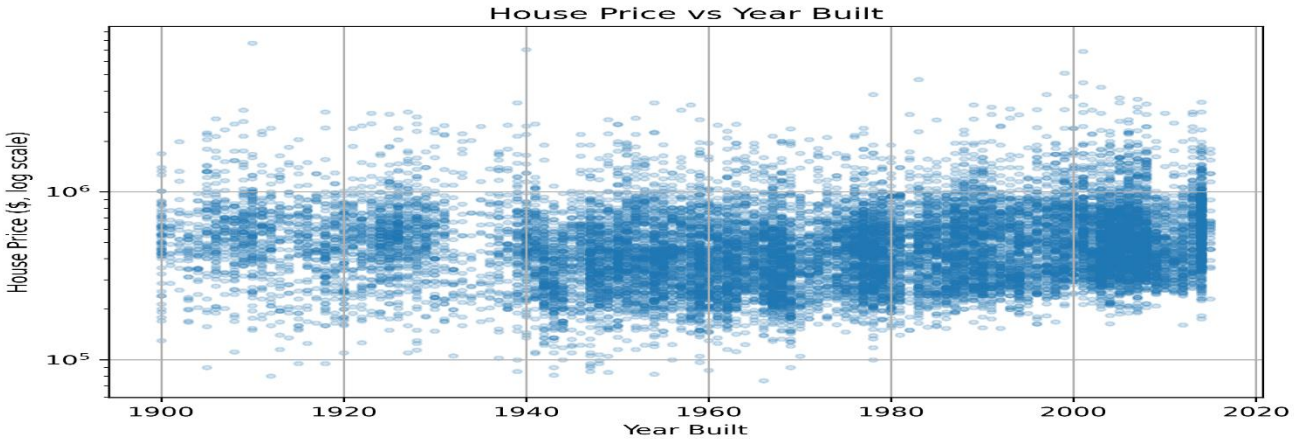
▪ Renovation Status vs House Price

- We analyze the impact of renovation using a box plot comparing renovated and non-renovated houses on a log-price scale. The plot shows a **clear upward shift in the median price for renovated properties**, along with a higher upper quartile. This indicates that renovation contributes a **distinct price premium**, rather than merely increasing variance. However, the overlap between the two distributions suggests that renovation alone does not fully determine price—its effect interacts with other factors such as location, size, and grade. This observation motivates explicitly including renovation as a binary feature in the model rather than relying on age or size to implicitly capture its effect.



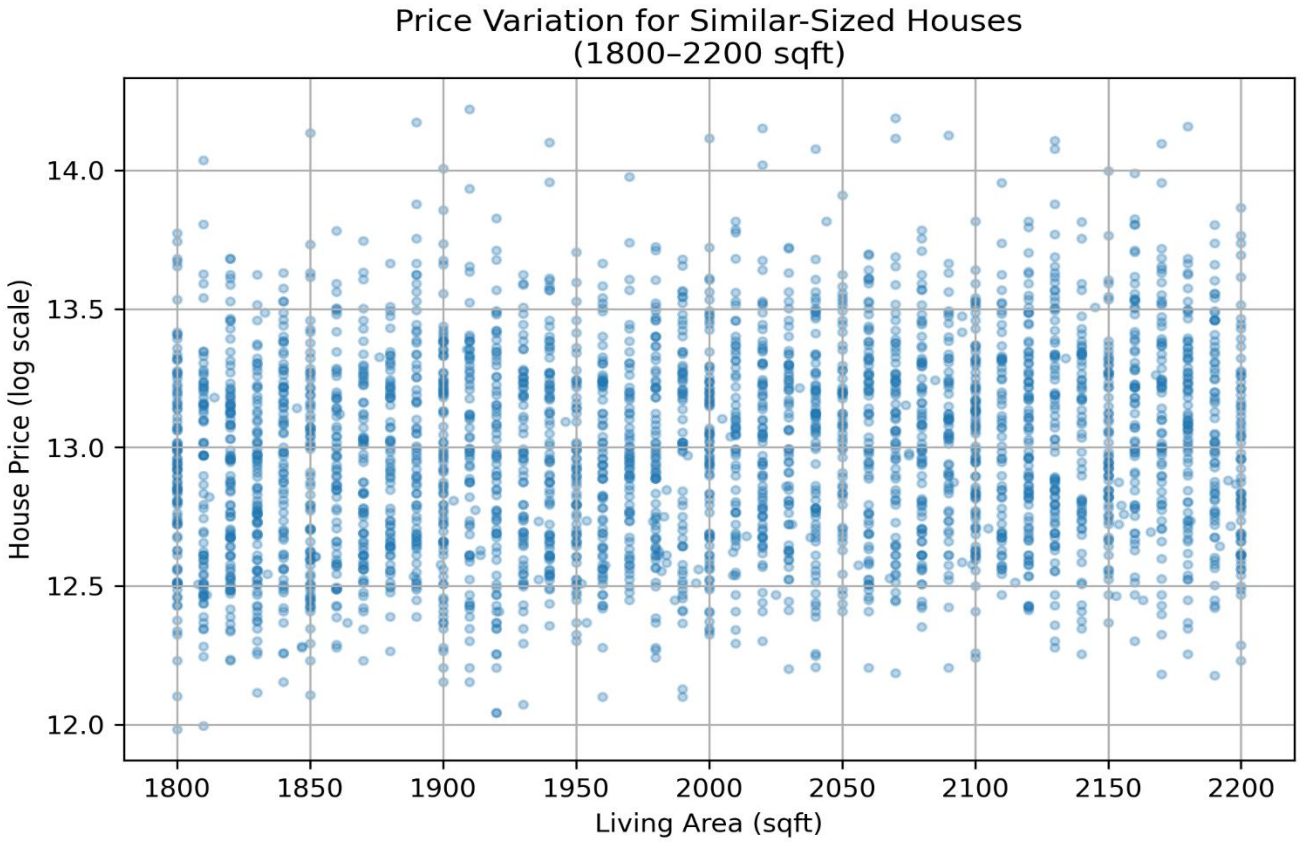
Year Built vs House Price

The scatter plot of house price against year built reveals a **weak but noticeable upward trend**, where newer houses tend to have higher prices on average. However, the large vertical spread for each construction year indicates **substantial price variability among houses built in the same period**. This suggests that while property age carries useful signal, it is **not a dominant predictor by itself**. The pattern highlights the limitation of linear age-based assumptions and justifies using models capable of learning **non-linear interactions between year built, renovation status, and other structural or locational features**.



5. Price Variability for Similar-Sized Houses

- To isolate the effect of non-size-related factors, we analyze houses with **similar living area (1800–2200 sqft)** and examine how their prices vary. Since living area is one of the strongest predictors of price, restricting the analysis to a narrow size range allows us to control for this dominant factor and study residual price differences.
- Despite the near-constant living area, the plot shows **substantial spread in house prices (log scale)**. Homes of almost identical size exhibit price differences spanning **multiple orders of magnitude**, indicating that size alone is insufficient to explain valuation. This variability persists across the entire selected range and is not attributable to measurement noise.
- This dispersion suggests the presence of **latent factors not captured by basic structural attributes**, such as neighborhood quality, proximity to water or greenery, road connectivity, and surrounding urban layout. These factors are inherently spatial and visual in nature, motivating the incorporation of **satellite imagery** into the modeling pipeline. This observation directly supports the need for a **hybrid model** that combines tabular features with visual context extracted via a CNN.



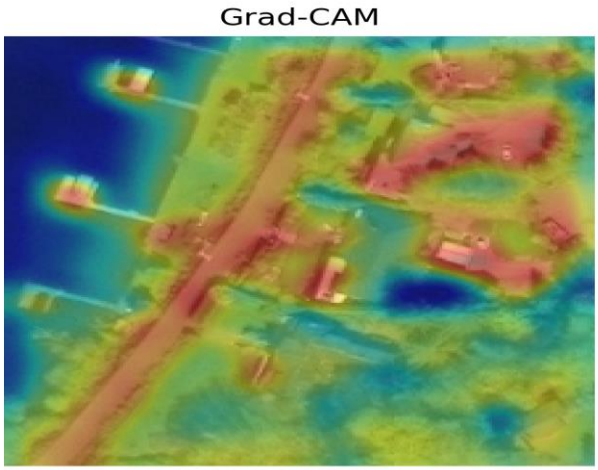
Houses with nearly identical size can have significantly different prices, implying that location and visual environment play a critical role in valuation.

Visual Feature Analysis Using Grad-CAM

To understand how satellite imagery contributes to price prediction, we analyze Grad-CAM visualizations generated from the trained CNN component of the hybrid model. Grad-CAM highlights image regions that most strongly influence the predicted price, allowing us to identify **visual cues the model has internally learned**, rather than manually defined features.

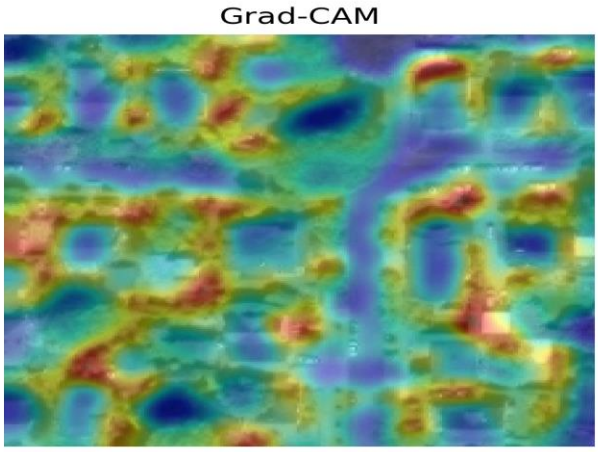
Road Structure and Spatial Organization

In the first image, Grad-CAM activations are strongly aligned with **structured road networks and intersections**, rather than individual rooftops alone. The model assigns high importance to continuous, well-organized road geometry, indicating that it has learned to associate **urban layout and accessibility patterns** with higher prices. This suggests the CNN captures neighborhood-level spatial organization as a pricing signal, beyond isolated property characteristics.



Joint Attention to Built Structures and Green Cover

The second heatmap shows distributed activation across **building rooftops, internal roads, and surrounding greenery**. Instead of focusing on a single object, the model attends to the **coexistence of dense construction and vegetative cover**. This indicates that the CNN has learned a composite visual representation of neighborhood quality, where balanced land use and environmental presence jointly contribute to higher price predictions.

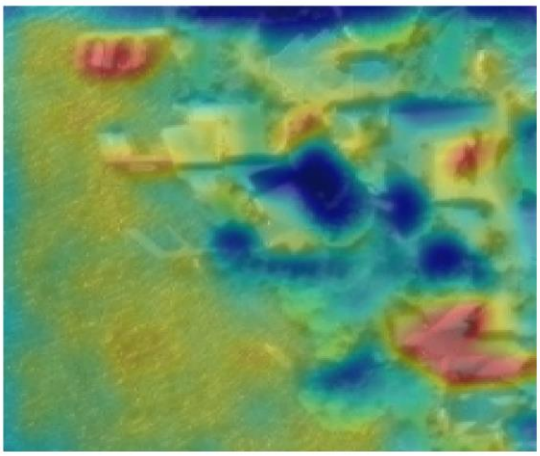


Waterfront Proximity as a Dominant Visual Cue

- In the third heatmap, Grad-CAM exhibits the strongest activation along the **water boundary**, dominating other visual elements. This confirms that the model explicitly recognizes **proximity to water bodies** as a high-impact pricing factor. Importantly, this activation emerges directly from pixel-level learning, without explicit geographic features, demonstrating that the CNN independently encodes waterfront presence as a premium signal.

Original Image

Grad-CAM

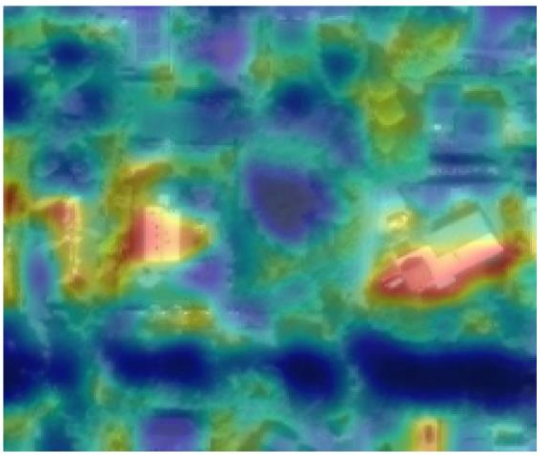


Rooftop Scale and Structural Prominence

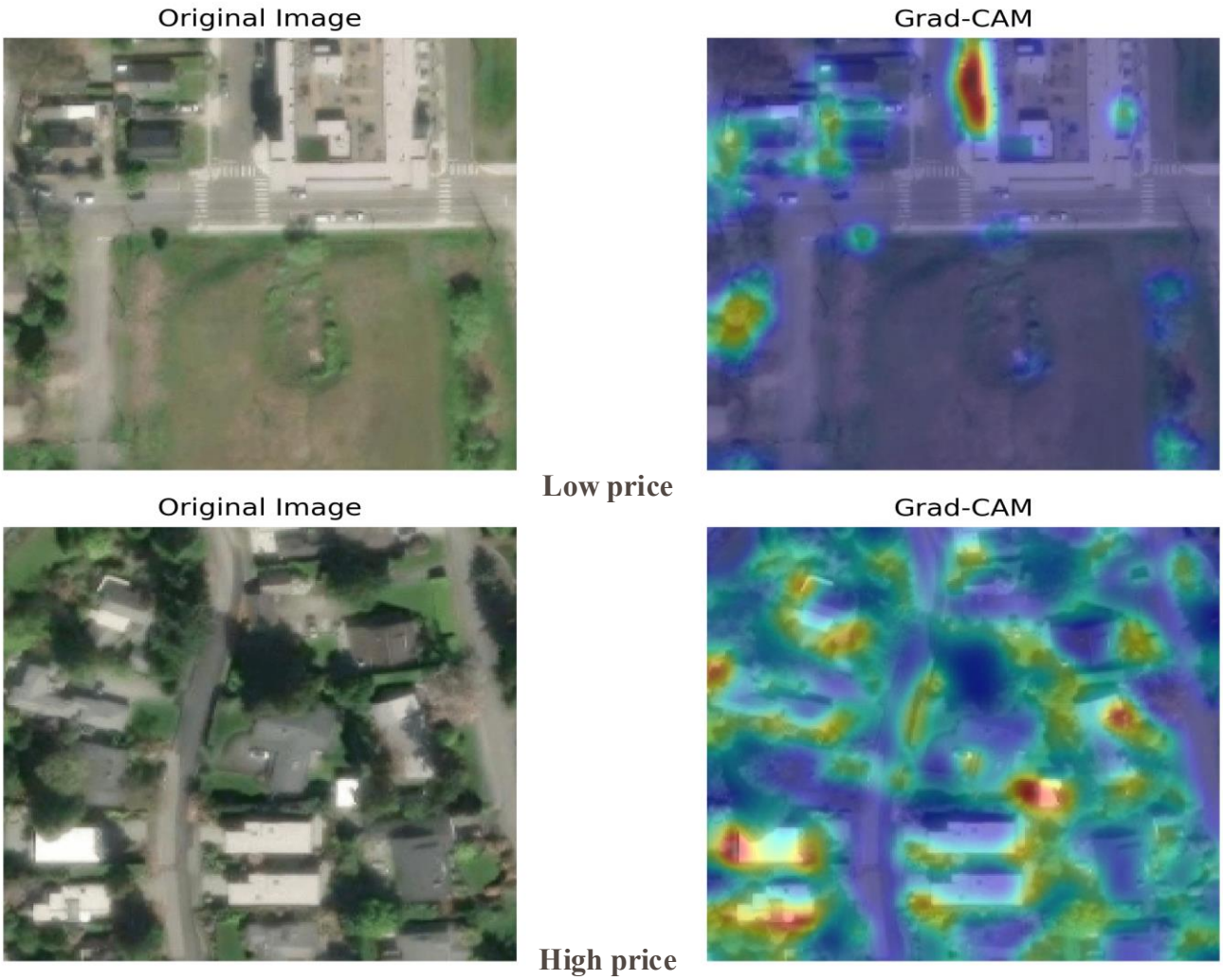
The fourth heatmap highlights concentrated activation over **large, clearly defined rooftops**, while surrounding regions receive comparatively lower attention. This suggests the model uses rooftop size and structural clarity as visual proxies for **building scale, construction quality, and property prominence**. Such features likely correlate with larger living areas and higher-grade construction, reinforcing their influence on price estimation.

Original Image

Grad-CAM



Green Cover vs Built Density

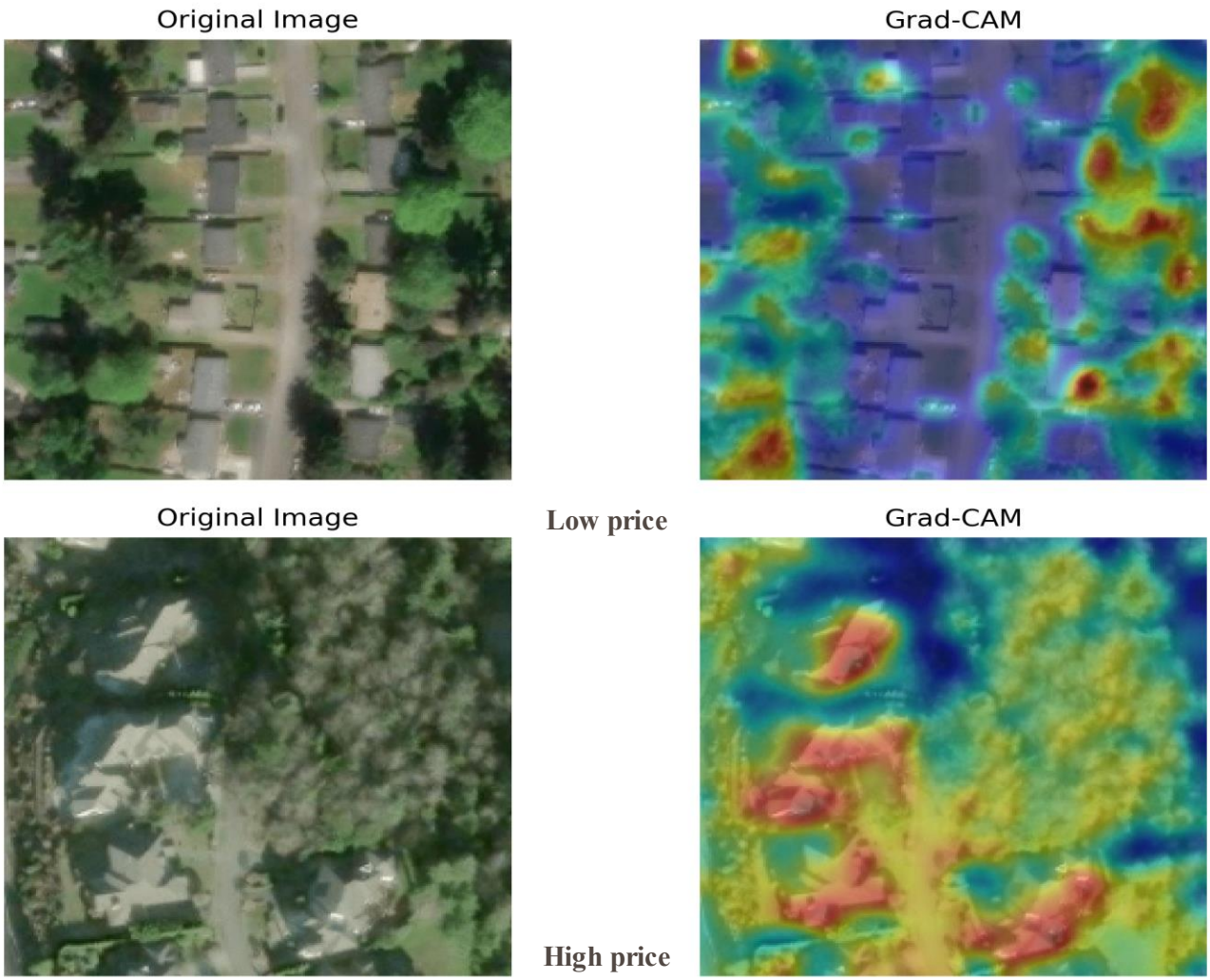


In the **low-price image**, the original satellite view shows a large open green area with sparse surrounding construction. The corresponding Grad-CAM heatmap exhibits **weak and fragmented activations**, primarily concentrated near isolated structures and road intersections, while the majority of the green region remains low-activation (blue). This indicates that, for low-priced properties, extensive green cover **without structured development** contributes limited predictive signal to the model. Open land and unorganized greenery appear to be weak indicators of higher valuation.

In contrast, the **high-price image** displays a dense residential layout interwoven with greenery. The Grad-CAM visualization reveals **strong, spatially coherent activations** concentrated over **building footprints, rooftops, and road-connected blocks**, with secondary attention on surrounding vegetation. This suggests that the model associates **green cover embedded within a dense, well-planned built environment** as a strong positive pricing signal. Rather than raw greenery alone, the **integration of vegetation with structured housing density** is what the model learns as value-enhancing.

Overall, this comparison demonstrates that the model does not treat green cover as an isolated positive feature. Instead, Grad-CAM evidence shows that **greenery amplifies price only when coupled with organized built density**, reflecting real-world urban valuation patterns. These observations justify the inclusion of satellite imagery, as such spatial relationships cannot be captured through tabular features alone.

Prominent Rooftops vs Fragmented Structures



In the **low-priced house,** the Grad-CAM activations are relatively **diffuse and weak**, with attention spread irregularly across small rooftops, surrounding vegetation, and road segments. No single structure dominates the activation map. This fragmented pattern suggests that the model does not identify a strong, coherent visual signal associated with high value. The built environment appears dense but uniform, with limited architectural distinction, leading to weaker price cues from visual features alone.

In contrast, the **high-priced house** exhibits **strong, concentrated red activations** sharply localized over **large, well-defined rooftops and structured layouts**. The model consistently focuses on prominent roof geometries and organized spatial patterns, indicating that these features play a significant role in driving higher price predictions. The sharp activation boundaries imply that the model has learned to associate **architectural prominence, scale, and structural clarity** with increased property value.

Overall, this comparison demonstrates that the model distinguishes not merely the presence of buildings, but the **quality and organization of built structures**. Prominent, clearly defined rooftops act as strong positive visual signals, while fragmented or homogeneous structures contribute weakly, reinforcing the importance of architectural form in visual price estimation.

Model Architecture and Explainability Strategy

---

1. Hybrid Model Architecture for Price Prediction

- The final prediction system is implemented as a **hybrid multimodal neural network** that integrates **satellite imagery** with **structured tabular housing attributes**. This design is motivated by the observation that house prices are influenced not only by intrinsic property characteristics (e.g., size, number of rooms, construction year) but also by **neighborhood-level spatial context**, which cannot be reliably captured using tabular features alone.

- Formally, the hybrid model consists of two parallel processing branches:
  - A **convolutional neural network (CNN)** that extracts visual and spatial representations from satellite images.
  - A **fully connected multilayer perceptron (MLP)** that encodes standardized numerical property attributes.

- The representations learned by these branches are fused at the **feature level** and jointly optimized to predict the **log-transformed house price**.

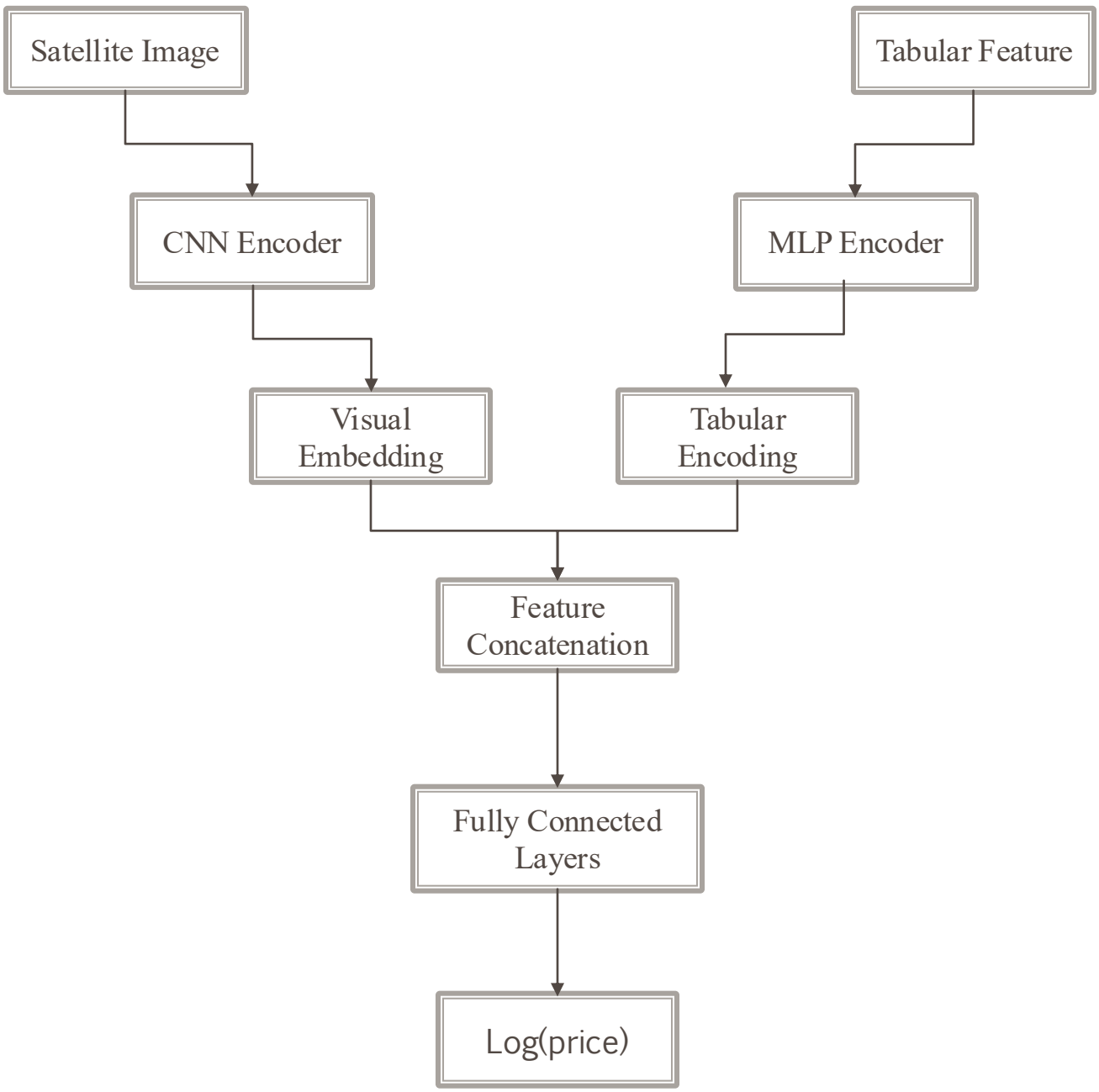
2. Image Branch in the Hybrid Model

- The image branch uses **EfficientNet-B0** as a convolutional feature extractor.
  - **2.1 Input Representation**
    - Input: RGB satellite image of resolution **224 × 224 × 3**
    - Pixel values normalized to $[0, 1]$
    - Images loaded dynamically using a **tf.data** pipeline for scalability

  - **2.2 Backbone and Fine-Tuning Strategy**
  - EfficientNet-B0 is initialized with **ImageNet-pretrained weights** and used without the classification head **(include_top=False)** To balance generalization and domain adaptation:
  - Early convolutional layers are **frozen** to preserve generic visual representations.
  - The **last 40 convolutional layers** are fine-tuned, allowing the model to adapt to housing-specific spatial patterns such as:
    - Road connectivity
    - Building layouts
    - Rooftop geometry
    - Green cover distribution
    - Waterfront proximity

  - **2.3 Visual Embedding Construction**
  - The CNN output feature maps are aggregated using **Global Average Pooling (GAP)**, followed by a projection head:
  - Dense (256, ReLU)
  - Dense (128, ReLU)
  - The result is a **128-dimensional visual embedding** that encodes neighborhood-level context. This embedding does **not** directly predict price; instead, it is fused with tabular features in the next stage.

- **3. Tabular Branch in the Hybrid Model**
  - The tabular branch processes **16 standardized numerical features**, including:
  - Structural attributes (bedrooms, bathrooms, floors)
  - Area-based features (living area, lot size, basement size)
  - Construction year
  - Geographic coordinates (latitude, longitude)
  - Neighborhood proxies (sqft_living15, sqft_lot15)
  - Binary indicators (waterfront, view)
  - All tabular features are standardized using **training-only statistics** to avoid data leakage.
  - The tabular encoder is implemented as a lightweight MLP:
  - Dense (128, ReLU)
  - Dense (64, ReLU)



- **4. Feature Fusion and Joint Prediction**
  - The outputs of the image and tabular branches are concatenated to form a joint feature representation:

$$z = [z_{image} \parallel z_{tabular}]$$

  - This fused representation is passed through additional fully connected layers:
  - Dense (128, ReLU)
  - Dense (64, ReLU)
  - The final output layer predicts a **single scalar corresponding to normalized log(price)**.
  - This fusion allows the model to learn **cross-modal interactions**, enabling visual context to modulate the influence of tabular attributes and vice versa.

- **5. Loss Function and Optimization Strategy**
  - House prices are modeled in **log space** using log1p(price) to reduce skewness and stabilize gradients. For the hybrid model:
  - Optimizer: **Adam**
  - Learning rate: 1e-4
  - Gradient clipping **(clipnorm = 1.0)**
  - Loss function: **Huber loss**
  - Huber loss provides robustness to outliers while retaining smooth optimization behaviour.

6. Staged Training Strategy

To ensure stable convergence, training is performed in **two conceptual stages**:

Stage 1: Stabilizing Fusion Layers

- CNN backbone is largely frozen
- Only:
  - Tabular branch
  - Projection layers
  - Fusion layers

  are trained
- Allows the model to learn a stable mapping from combined embeddings to price

Stage 2: Fine-Tuning Visual Features

- Upper layers of the CNN are unfrozen
- Learning rate is reduced
- Visual features adapt to domain-specific cues such as:
  - Road layouts
  - Building density
  - Rooftop structures
  - Environmental context
- This staged approach avoids catastrophic forgetting and reduces overfitting.

6. Why a Separate CNN-Only Model Is Required for Grad-CAM

- Although the hybrid model produces the final price predictions, **Grad-CAM cannot be applied to the hybrid architecture** in a theoretically valid manner.
- Grad-CAM requires a prediction that depends **solely on convolutional feature maps** to preserve spatial correspondence between pixels and output gradients. In the hybrid model, visual features are concatenated with tabular embeddings, breaking this spatial alignment.
- Applying Grad-CAM after feature fusion would therefore produce **misleading or uninterpretable explanations**.
- To address this, a **separate CNN-only regression model** is trained exclusively on satellite images and used **only for visual explainability**.

7. CNN-Only Model Architecture for Heatmap Generation

- The CNN-only model isolates the **visual reasoning component** of the system.
- **7.1 Architecture**
  - The CNN-only model consists of:
  - Input: 224 × 224 × 3 satellite image
  - Convolutional blocks:
    - Conv(32) → MaxPool
    - Conv(64) → MaxPool (**designated Grad-CAM layer: gradcam_conv**)
    - Conv(128)
  - Global Average Pooling
  - Dense (64, ReLU)
  - Dense (1, Linear) → normalized log(price)
  - This architecture is intentionally lightweight to ensure:
  - Stable training on large image datasets
  - Clear spatial gradients for Grad-CAM

8. Target Normalization and Training Stability (CNN-Only Model)

- To prevent training instability and exploding losses when scaling to **16k images**, the CNN-only model predicts a **normalized log-price**:

$$y_{norm} = \frac{\log(1 + price) - \mu}{\sigma}$$

- Where $\mu$and $\sigma$are computed **only on the training split**.
- Additional stability measures include:
- Gradient clipping (clipnorm = 1.0)
- Learning rate scheduling (ReduceLROnPlateau)
- Early stopping
- Termination on NaN detection

Results and Model Comparison

- This section presents the quantitative performance of the proposed models and evaluates the contribution of satellite imagery when combined with structured housing attributes. We compare a **Tabular-only baseline** against the proposed **Hybrid (Tabular + Satellite Image) model** using consistent data splits and evaluation metrics.

Evaluation Setup

- All models were evaluated on the same held-out validation set using a fixed train–validation split. House prices were modeled in **log-transformed space** to mitigate skewness and reduce the influence of extreme outliers. Model performance was assessed using **Root Mean Squared Error (RMSE)** and **coefficient of determination (R^2)** in log space.

Model	R^2 (log price)	Log RMSE	Approx % Error	Approx Dollar Error
Tabular-only	0.7129	0.281	32.5%	**$146,250**
Hybrid (Tabular + Images)	**0.8636**	**0.194**	**21.4%**	**$96,300**

- Although model evaluation was performed in log-price space, the results can be approximately interpreted in monetary terms. The tabular-only model achieves a log-RMSE of 0.281, corresponding to an average multiplicative error of approximately 32.5%. In contrast, the hybrid model incorporating satellite imagery reduces the log-RMSE to 0.194, corresponding to an average error of approximately 21.4%. At the dataset median house price of $450,000, this translates to an approximate reduction in absolute prediction error from $146,250 to $96,300, representing an improvement of nearly $50,000 per prediction. This demonstrates the substantial practical value of visual environmental information in real estate price estimation.

Performance Analysis

- The **Tabular-only model** achieves a reasonably strong baseline performance, indicating that structured attributes such as living area, location coordinates, grade, and condition explain a significant fraction of house price variability. However, its error remains relatively high, suggesting that important contextual information is not captured by tabular features alone.
- The **Hybrid model**, which integrates satellite imagery through a convolutional neural network alongside tabular inputs, demonstrates a **substantial improvement** over the baseline. The RMSE in log space decreases from **0.2815 to 0.1940**, corresponding to a relative error reduction of approximately **31%**. Simultaneously, the R^2 score increases from **0.713 to 0.864**, indicating that the hybrid approach explains a significantly larger proportion of variance in house prices.

Interpretation of Results

- The observed performance gain confirms that **visual environmental context provides complementary information** beyond traditional housing attributes. Satellite imagery implicitly captures factors such as neighborhood layout, green cover, road structure, building density, and proximity to water bodies—signals that are either absent or only weakly approximated in tabular data.
- By jointly learning from **explicit numerical features** and **implicit visual cues**, the hybrid architecture is able to form a richer representation of property value drivers. This result empirically validates the core hypothesis of this project: **house prices are influenced not only by intrinsic property characteristics but also by the surrounding visual environment**.

Summary

- Overall, the results demonstrate that augmenting tabular real estate data with satellite imagery leads to **significant and consistent performance gains**. The hybrid model substantially outperforms the tabular-only baseline, highlighting the effectiveness of multimodal learning for real-world price prediction tasks and motivating the architectural choices adopted in this work.