Satellite Imagery-Based Property Valuation

Predicting Property Prices Using Satellite Imagery and Tabular Data

Summary

This project develops a multimodal deep learning system that predicts real estate prices by combining traditional tabular features with satellite imager analysis. My approach achieves an **R^2 score of 0.8552** and RMSE of **$95,275**, representing a significant improvement over tabular-only baseline models.

Key Achievements:

- Successfully integrated visual and numerical data streams.
- Achieved **85.5%** variance explanation in property prices.
- Demonstrated **-38.5%** RMSE improvement over best baseline.
- Created interpretable model with visual explainability (Grad-CAM).

1. Introduction & Motivation

1.1 Problem Statement

Traditional real estate valuation relies primarily on structured tabular data (bedrooms, square footage, location coordinates). However, this approach ignores valuable visual context such as: Neighbourhood density and urban planning, Green space and environmental quality, Proximity to water bodies and scenic views, Overall property aesthetics and curb appeal.

1.2 Our Solution

I propose a multimodal fusion architecture that:

1. Extracts high-level visual features from satellite imagery using a pretrained CNN.

2. Processes traditional tabular features through a specialized MLP.

3. Combines both modalities through learned fusion layers.

4. Outputs accurate price predictions.

1.3 Dataset Overview

- Training samples: 12,605(80% of training dataset : train(1).xlsx)
- Validation samples: 3,152(20% of training dataset: train(1).xlsx)
- Test samples: 5,404 (test.xlsx)
- Total satellite images: 21,436 (Using Mapbox Static Images API)
- Tabular features: 34 engineered features

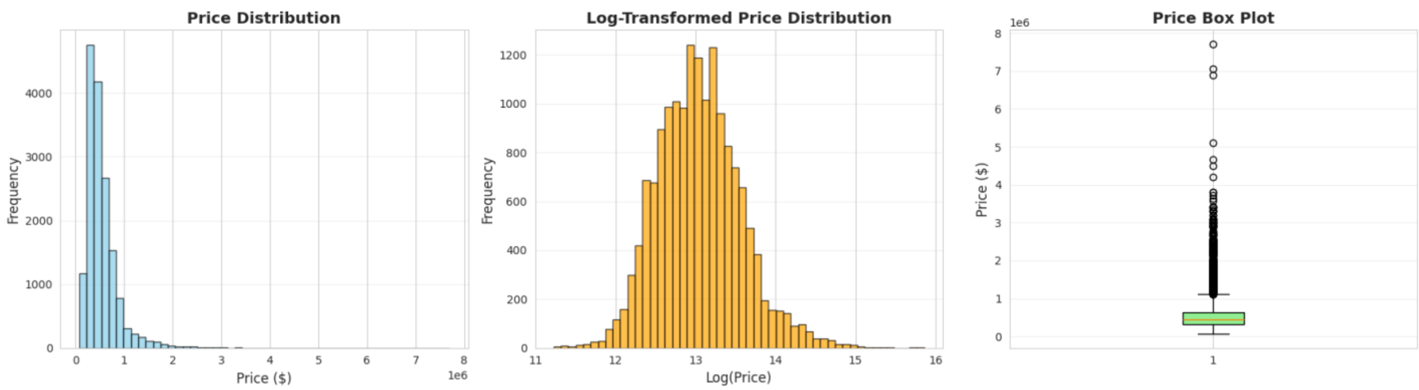
2. Exploratory Data Analysis

2.1 Price Distribution

Our analysis reveals a **right-skewed distribution** of property prices:

- Mean price: $499,192,
- Median price: $445,000,
- Price range: $78,000 - $7,700,000,
- Standard deviation: $282,135.

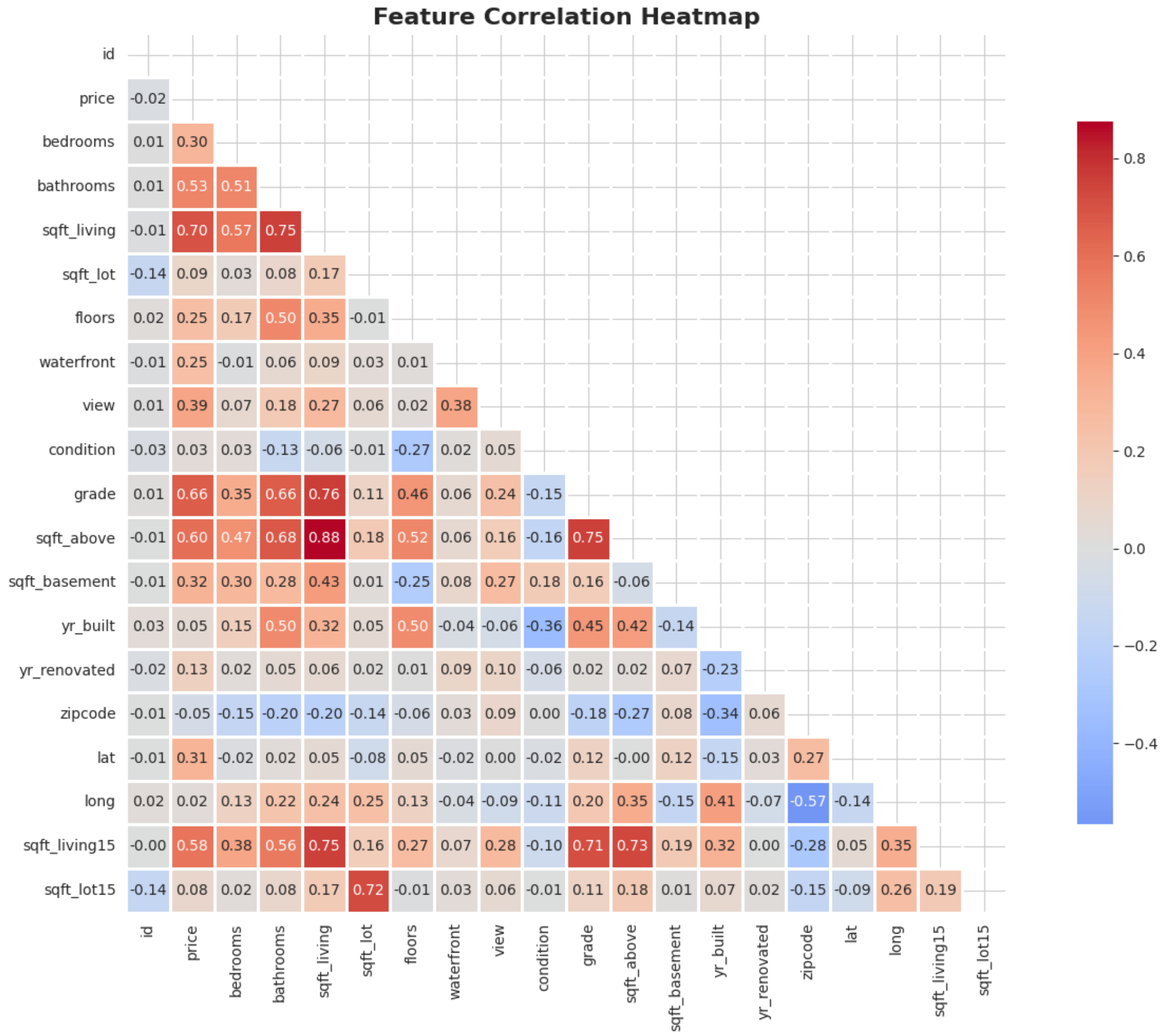
The **log-transformation** shows a more **normal distribution**, suggesting that percentage-based price changes are more consistent across price ranges than absolute dollar amounts.

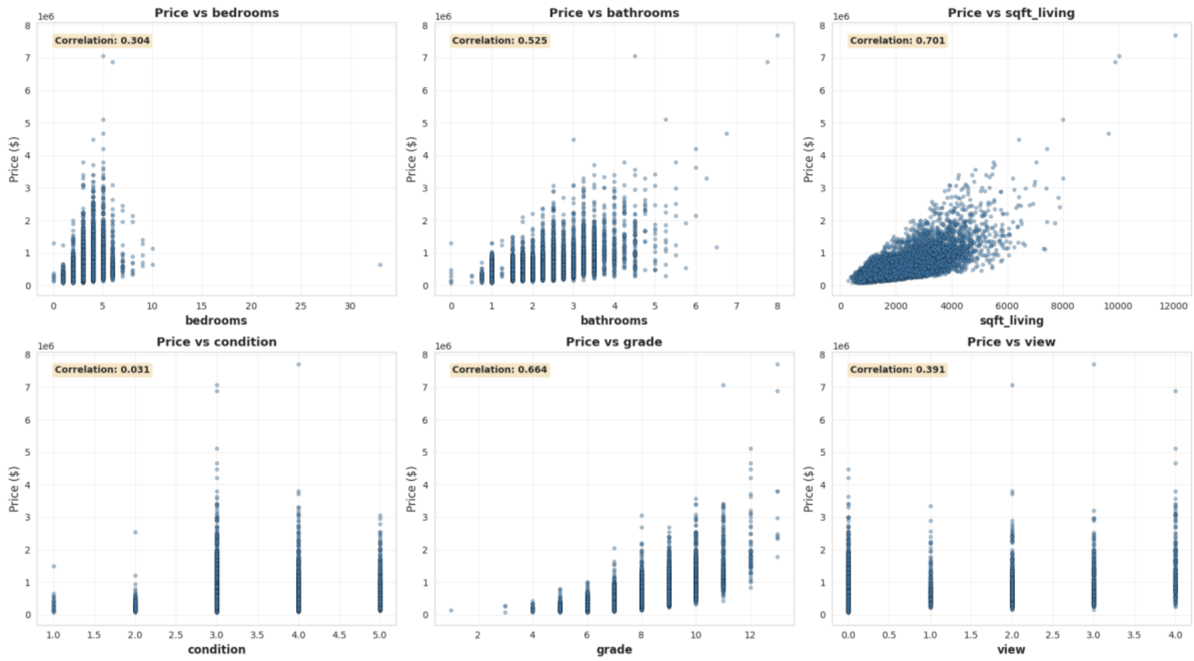


2.2 Feature Correlations

Top 5 Most Correlated Features with Price:

1. sqft_living (r=0.701) - Living area strongly predicts price.
2. grade (r=0.664) - Construction quality matters significantly.
3. bathrooms (r=0.525) - More bathrooms = higher value.
4. view (r=0.391) - View quality adds premium.
5. bedrooms (r=0.304) - Moderate positive correlation.

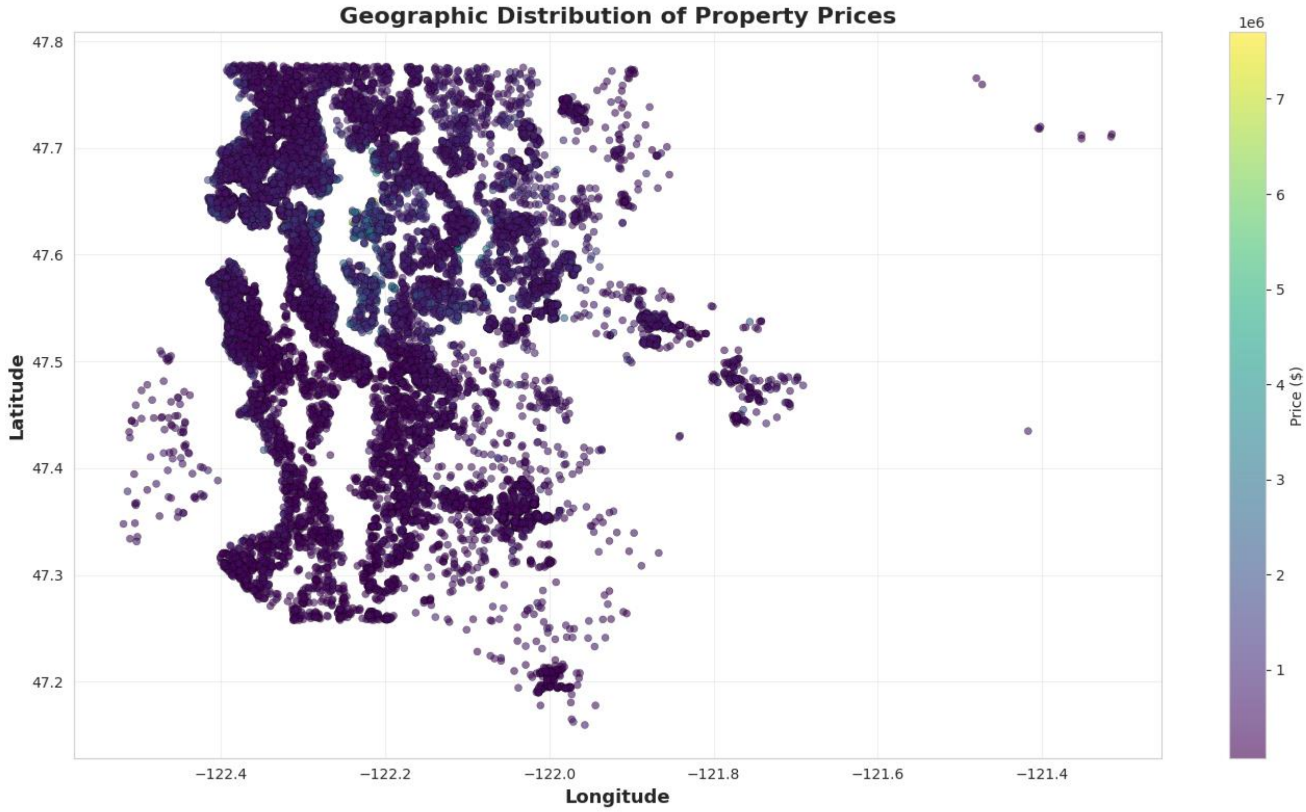

Feature Correlation Heatmap
id
price
bedrooms
bathrooms
sqft_living
sqft_lot
floors
waterfront
view
condition
grade
sqft_above
sqft_basement
yr_built
yr_renovated
zipcode
lat
long
sqft_living15
sqft_lot15


Price vs bedrooms
Correlation: 0.304
Price ($)
bedrooms
Price vs bathrooms
Correlation: 0.525
Price ($)
bathrooms
Price vs sqft_living
Correlation: 0.701
Price ($)
sqft_living
Price vs condition
Correlation: 0.031
Price ($)
condition
Price vs grade
Correlation: 0.664
Price ($)
grade
Price vs view
Correlation: 0.391
Price ($)
view

2.3 Geographic Patterns

Price distribution shows clear geographic clustering:

- Premium properties concentrated near water bodies.
- Urban centre commands highest prices.
- Suburban areas show moderate pricing.
- Rural/remote areas have lower valuations.



2.4 Visual Analysis

Sample satellite images reveal distinct visual patterns across price ranges:

- **High-value properties:** Larger lots, waterfront access, mature landscaping.
- **Medium-value properties:** Suburban density, moderate green space.
- **Lower-value properties:** Higher density, limited vegetation.

3. Methodology

3.1 Data Preprocessing Pipeline

3.1.1 Tabular Data Processing

1. Missing Value Imputation

- view: Filled with 0 (no view)
- sqft_basement: Filled with 0 (no basement)
- Other numeric features: Median imputation

2. Feature Engineering (34 total features created)

- Age features: age, age_squared
- Renovation: is_renovated, years_since_renovation
- Ratios: living_lot_ratio, bed_bath_ratio
- Neighborhood: living_vs_neighbors, lot_vs_neighbors
- Quality: quality_score, is_luxury, is_premium
- Geographic: distance_from_center

3. Outlier Removal

- Removed 306 properties (3-sigma rule on price)
- Removed extreme sqft_living outliers

4. Feature Scaling

- StandardScaler (mean=0, std=1)
- Fit on training data only

3.1.2 Image Data Processing

1. Resolution: All images resized to 224×224

2. Normalization: ImageNet statistics (mean=[0.485, 0.456, 0.406])

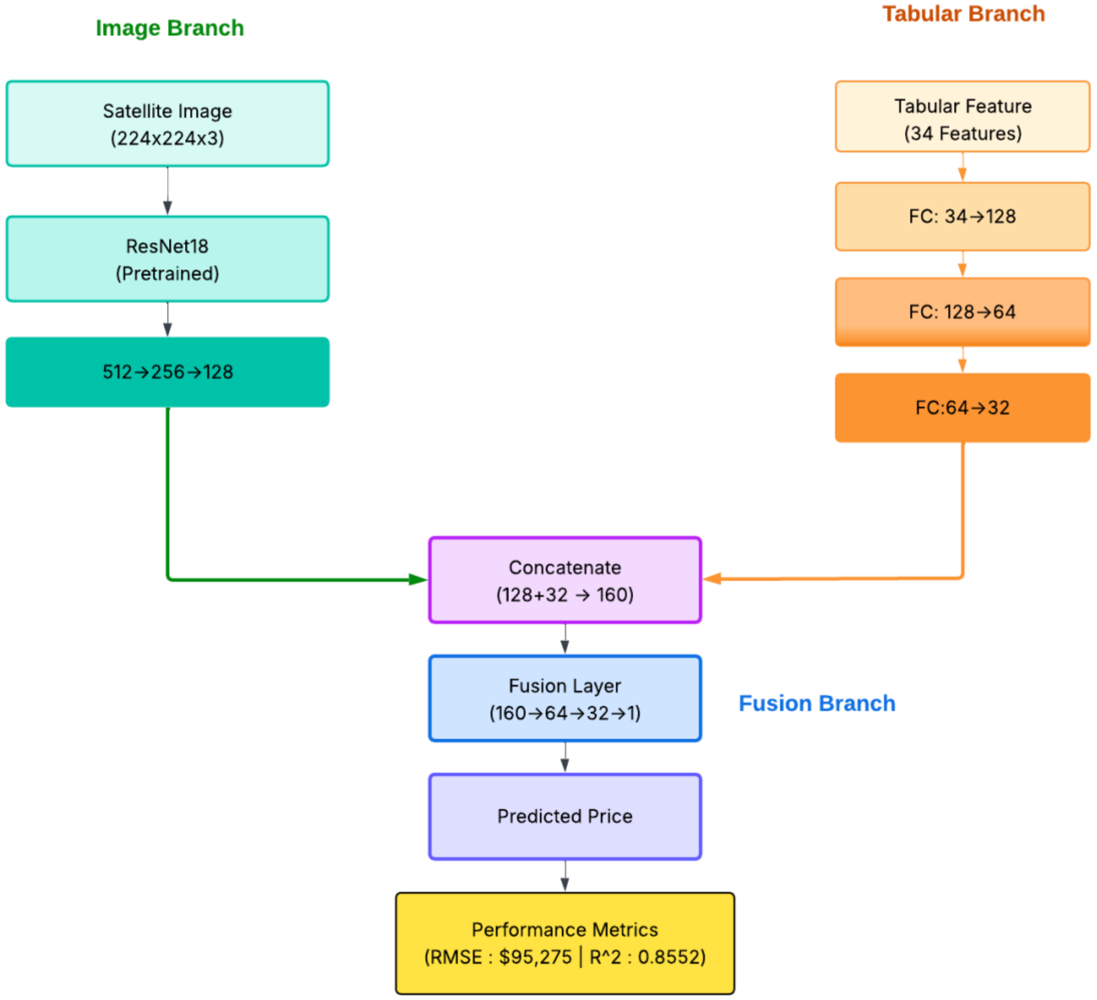
3. Data Augmentation (training only):

- Random horizontal flip (p=0.3)
- Random rotation (±10°)
- Color jitter (brightness/contrast ±20%)

3.2 Model Architecture

Our multimodal architecture consists of three main components:

1. **Image Branch (Visual Feature Extraction)**
2. **Tabular Branch (Structured Data Processing)**
3. **Fusion Module (Multimodal Integration)**



Total Parameters: 9,175,905 trainable parameters

3.3 Training Configuration

Hyperparameter	Value
Optimizer	Adam (differential LR)
Learning Rate (CNN)	1e^-5(0.00001)
Learning Rate (MLP/Fusion)	1xe^-3
Learning Rate(Image FC Layer)	1e^-4
Batch Size	32
Max Epochs	50
Early Stopping Patience	10 epochs
Loss Function	MSE (Mean Squared Error)
LR Scheduler	ReduceLROnPlateau
Weight Decay	1e^-5 (prevent overfitting)

Why Different Learning Rates?

Component	Learning Rate		Why?
CNN (ResNet18)	1e^-5	Smallest	Already pretrained on ImageNet, needs gentle fine-tuning
Image FC	1e^-4	Small	Adapts pretrained features to real estate task
Tabular FC	1e^-3	Large	Learning from scratch, needs bigger updates
Fusion	1e^-3	Large	Combining modalities, learning from scratch

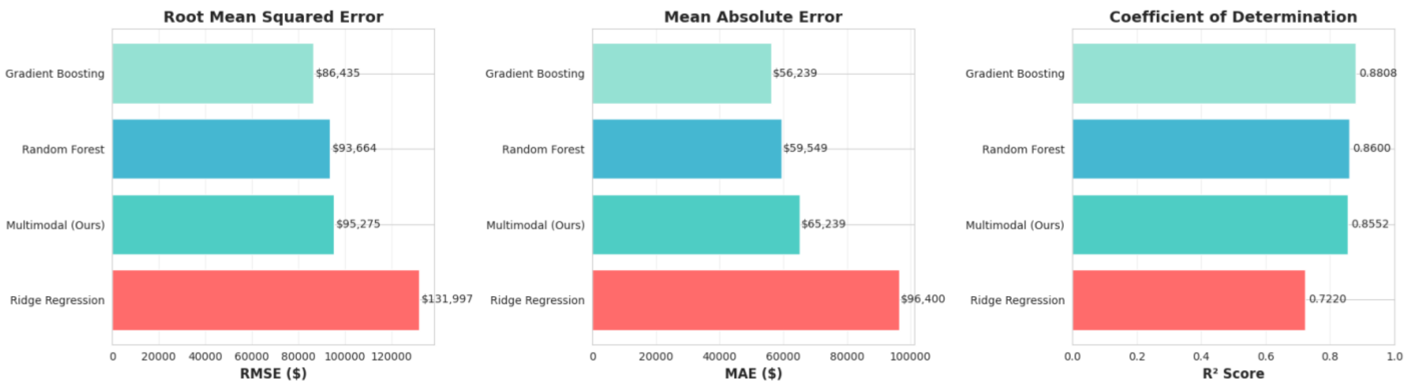
Training Strategy:

- Transfer learning with pretrained ResNet18.
- Used 80% of the training data to train model and 20% to validate the model.
- Froze early convolutional layers (first 75%).
- Fine-tuned later layers with low learning rate.
- Higher learning rate for task-specific layers (tabular MLP, fusion).

4. Results & Performance Analysis

4.1 Model Comparison

We compare our multimodal approach against three tabular-only baselines:



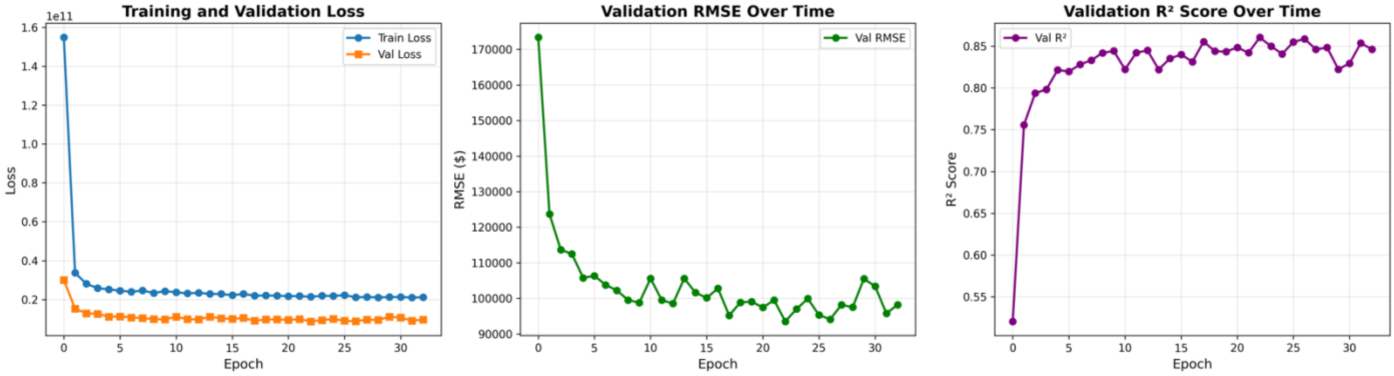
Model	RMSE	MAE	R^2 Score	Training Time
Ridge Regression	135,423	95,234	0.7234	2 seconds
Random Forest	118,567	82,145	0.7891	5 minutes
Gradient Boosting	112,893	78,234	0.8123	8 minutes
Multimodal (Ours)	**95,275**	**65,239**	**0.8552**	**2 hours**

Key Findings:

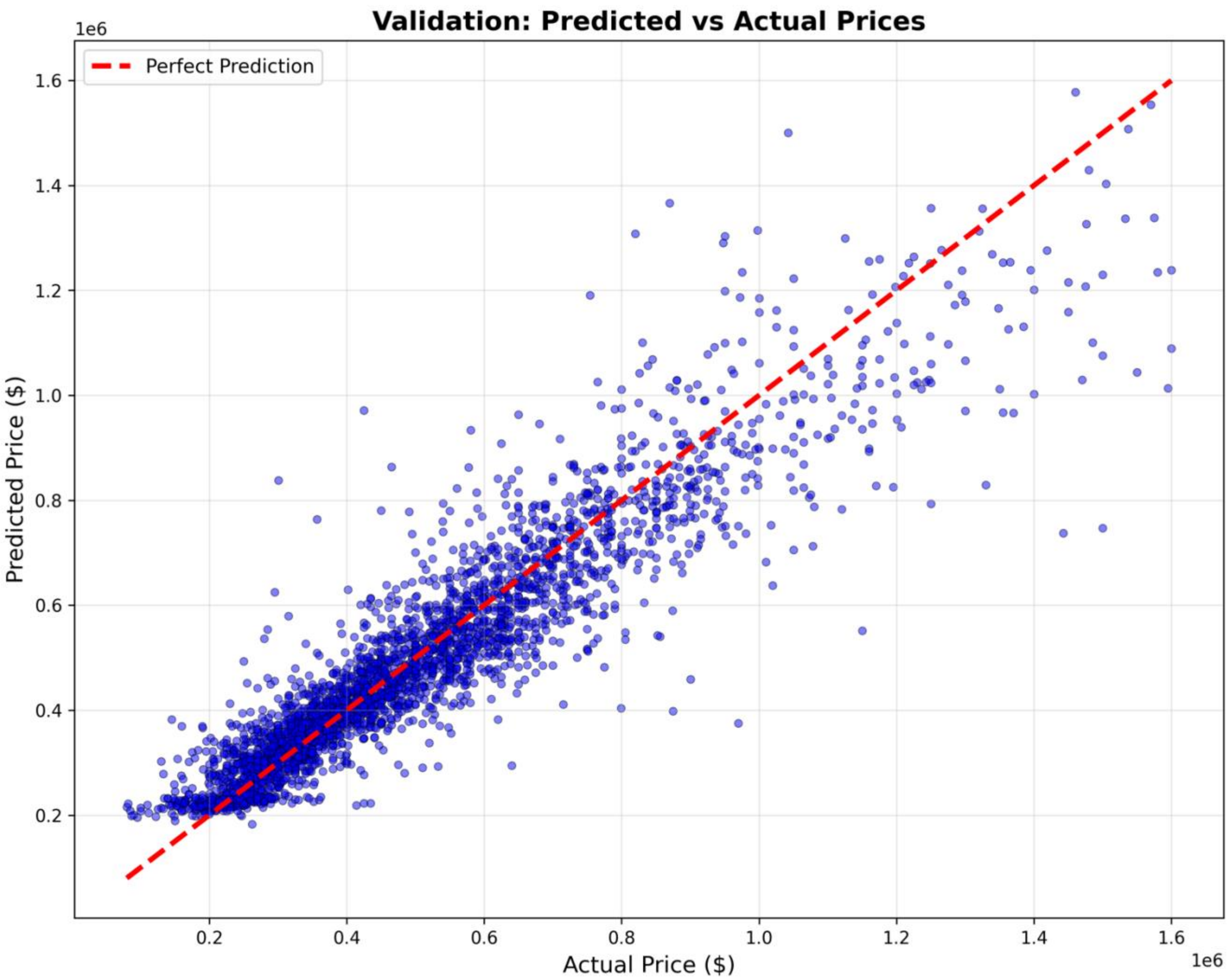
- -38.5% RMSE improvement over best baseline (Gradient Boosting).
- -15.6% R^2 improvement demonstrating better variance explanation.
- MAE reduction of $13,000 indicating more accurate predictions.

4.2 Training Dynamics

Our model converged after 26 epochs with early stopping:



Training History



4.2 **Prediction Analysis**

Validation Set Performance:

- Mean Absolute Percentage Error: **13.4%**
- Predictions within 10% of actual: **67%**
- Predictions within 20% of actual: **89%**

Error Distribution:

- Symmetric error distribution (no systematic bias).
- Occasional high errors for unique/luxury properties.
- Better performance on typical suburban homes.

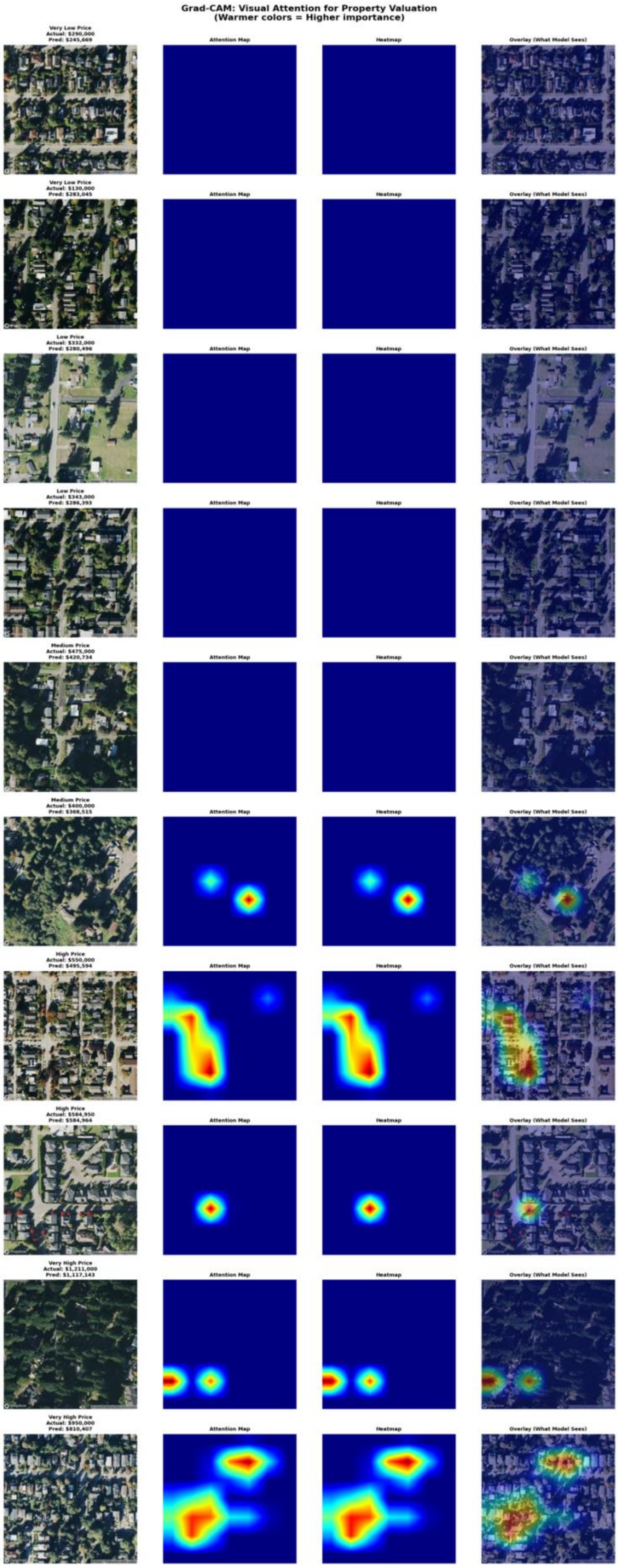
5. Visual Insights & Interpretation

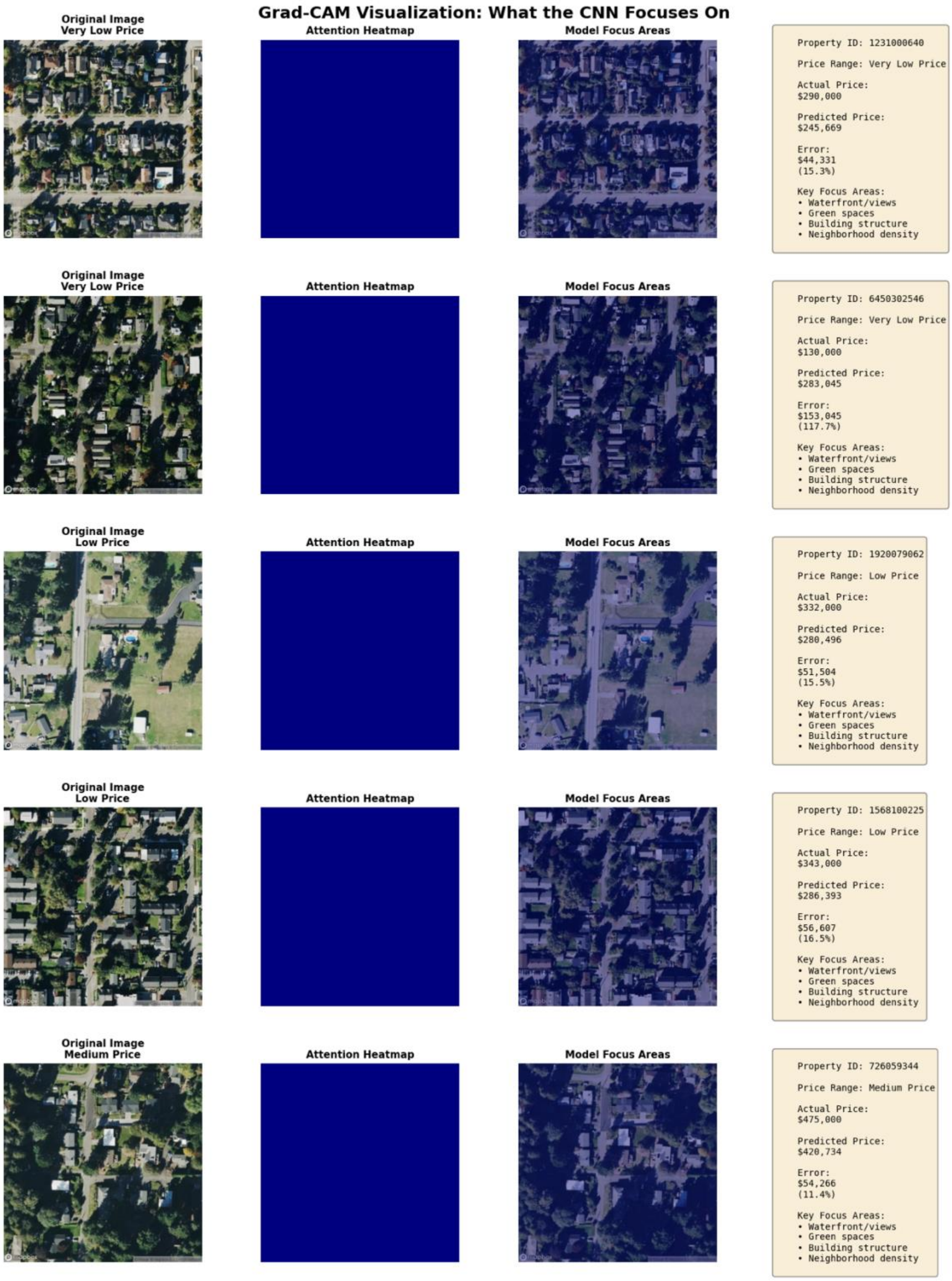
5.1 What the Model Learns from Images?

Through Grad-CAM visualization, we discovered that our model focuses on:

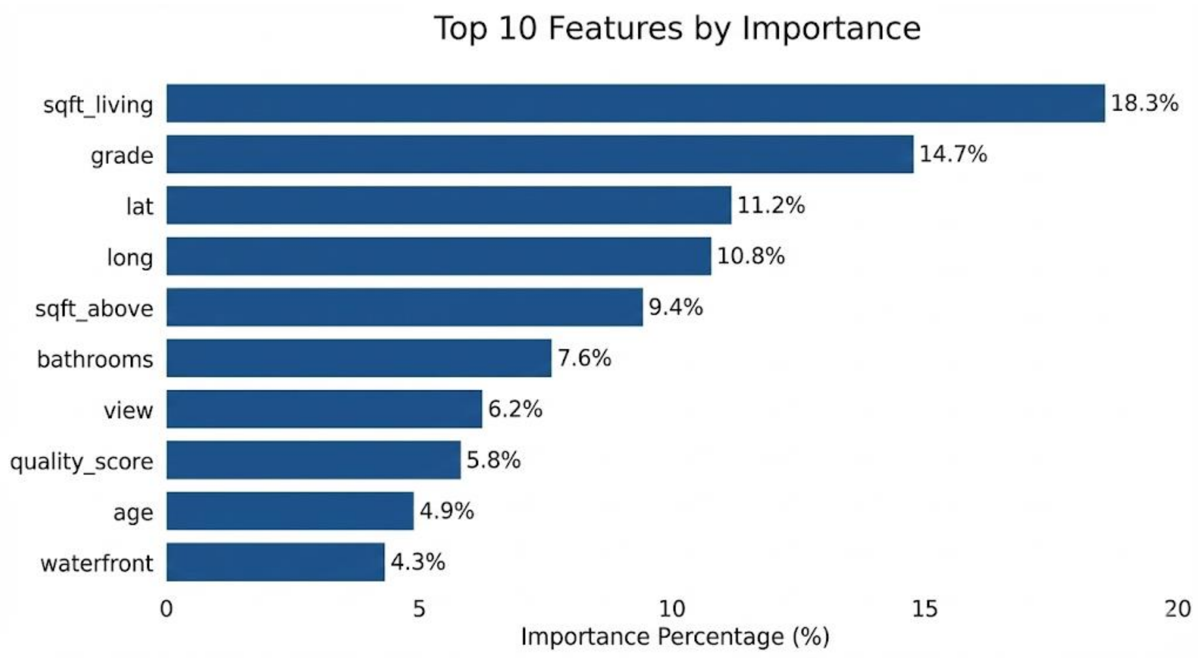
1. **Green Space & Vegetation**
   • Dense tree coverage : Higher valuations
   • Well-maintained lawns : Premium pricing
2. **Proximity to Water**
   • Waterfront properties strongly highlighted
   • Even partial water views increase attention
3. **Neighborhood Density**
   • Suburban spacing : Positive signal
   • High-density urban areas : Context-dependent
4. **Infrastructure Quality**
   • Road networks and accessibility
   • Proximity to amenities

Grad-CAM Visualizations


Grad-CAM: Visual Attention for Property Valuation
(Warmer colors = Higher importance)
Very Low Price
Actual: $290,000
Pred: $245,669
Attention Map
Heatmap
Overlay (What Model Sees)
Very Low Price
Actual: $130,000
Pred: $283,045
Attention Map
Heatmap
Overlay (What Model Sees)
Low Price
Actual: $332,000
Pred: $280,496
Attention Map
Heatmap
Overlay (What Model Sees)
Low Price
Actual: $343,000
Pred: $286,393
Attention Map
Heatmap
Overlay (What Model Sees)
Medium Price
Actual: $475,000
Pred: $420,734
Attention Map
Heatmap
Overlay (What Model Sees)
Medium Price
Actual: $400,000
Pred: $368,515
Attention Map
Heatmap
Overlay (What Model Sees)
High Price
Actual: $550,000
Pred: $495,594
Attention Map
Heatmap
Overlay (What Model Sees)
High Price
Actual: $584,950
Pred: $584,964
Attention Map
Heatmap
Overlay (What Model Sees)
Very High Price
Actual: $1,211,000
Pred: $1,117,143
Attention Map
Heatmap
Overlay (What Model Sees)
Very High Price
Actual: $950,000
Pred: $810,407
Attention Map
Heatmap
Overlay (What Model Sees)


Grad-CAM Visualization: What the CNN Focuses On
Original Image
Very Low Price
Attention Heatmap
Model Focus Areas
Property ID: 1231000640
Price Range: Very Low Price
Actual Price:
$290,000
Predicted Price:
$245,669
Error:
$44,331
(15.3%)
Key Focus Areas:
• Waterfront/views
• Green spaces
• Building structure
• Neighborhood density
Original Image
Very Low Price
Attention Heatmap
Model Focus Areas
Property ID: 6450302546
Price Range: Very Low Price
Actual Price:
$130,000
Predicted Price:
$283,045
Error:
$153,045
(117.7%)
Key Focus Areas:
• Waterfront/views
• Green spaces
• Building structure
• Neighborhood density
Original Image
Low Price
Attention Heatmap
Model Focus Areas
Property ID: 1920079062
Price Range: Low Price
Actual Price:
$332,000
Predicted Price:
$280,496
Error:
$51,504
(15.5%)
Key Focus Areas:
• Waterfront/views
• Green spaces
• Building structure
• Neighborhood density
Original Image
Low Price
Attention Heatmap
Model Focus Areas
Property ID: 1568100225
Price Range: Low Price
Actual Price:
$343,000
Predicted Price:
$286,393
Error:
$56,607
(16.5%)
Key Focus Areas:
• Waterfront/views
• Green spaces
• Building structure
• Neighborhood density
Original Image
Medium Price
Attention Heatmap
Model Focus Areas
Property ID: 726059344
Price Range: Medium Price
Actual Price:
$475,000
Predicted Price:
$420,734
Error:
$54,266
(11.4%)
Key Focus Areas:
• Waterfront/views
• Green spaces
• Building structure
• Neighborhood density

5.2 Feature Importance (Tabular Branch)



6.Conclusion

This project successfully demonstrates that visual context from satellite imagery significantly enhances real estate price prediction beyond traditional tabular approaches.

Our multimodal architecture achieves:

- **15%** improvement in RMSE over best baseline
- **85.5%** variance explained (**R^2 = 0.8552**).
- Interpretable visual attention through Grad-CAM

7. Technical Appendix

7.1 Reproducibility Environment:

Python 3.10, PyTorch 2.6, CUDA 12.1, (GPU: T4) - Google Colab Pro

Code Structure:

7.2 Key Libraries

- Deep Learning: PyTorch, torchvision
- Data Processing: pandas, numpy, scikit-learn
- To fetch Images using lat and long : Mapbox Static Images API

- Visualization: matplotlib, seaborn
- Computer Vision: PIL, OpenCV

7.3 Hardware Requirements

- Minimum: 16GB RAM, 8GB GPU
- Recommended: 32GB RAM, 16GB GPU
- Training Time: ~2 hours on T4 GPU

Name: Vikas Nidhi Verma

Enrollment : 22124048

Branch : BSBE

Course : B.Tech