Satellite Imagery–Based Property Valuation using Multimodal Deep Learning

Devanshi Jolhe

23117046

Problem Motivation

- **Why is property valuation difficult?**
- House prices depend on more than structural attributes
- Environmental & neighborhood context plays a critical role
- Traditional tabular models ignore visual surroundings
- **Goal of this project**
- Predict property prices by combining:
  - Structured property data
  - Satellite imagery from latitude & longitude

Dataset Overview

Tabular Dataset

- Source: Kaggle House Sales Dataset
- Key features:
  - Bedrooms, bathrooms, sqft_living
  - Condition, grade, view
  - Latitude & longitude

Target: **Price**

Visual Dataset

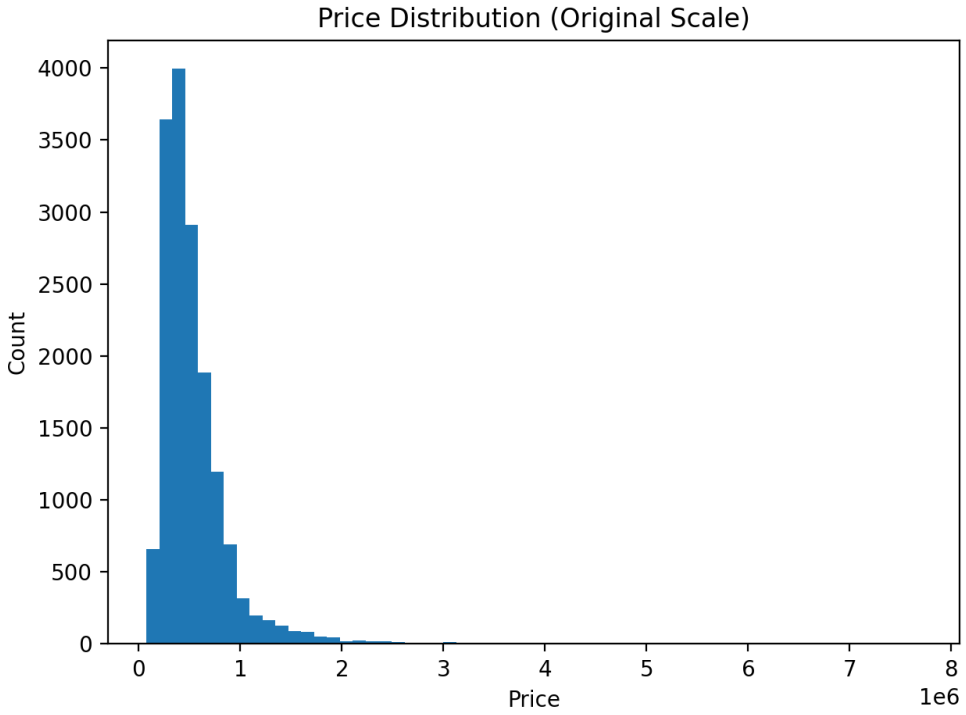
- Satellite images fetched using coordinates
- Captures greenery, roads, water bodies, density

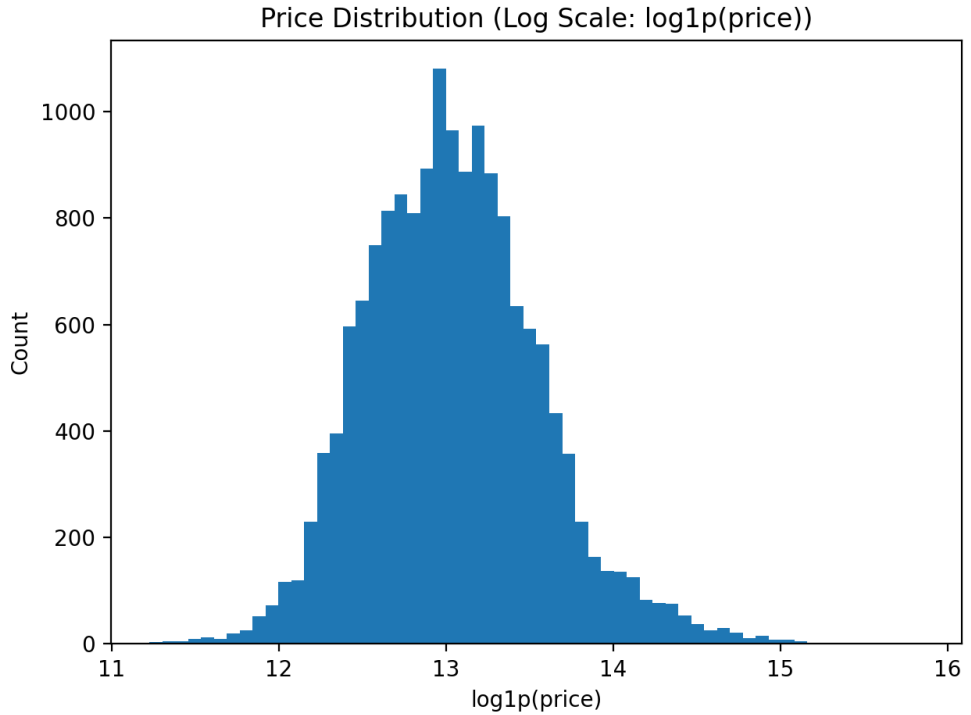
Sample satellite images from testing set

Exploratory Data Analysis (EDA): Price Distribution

Observations

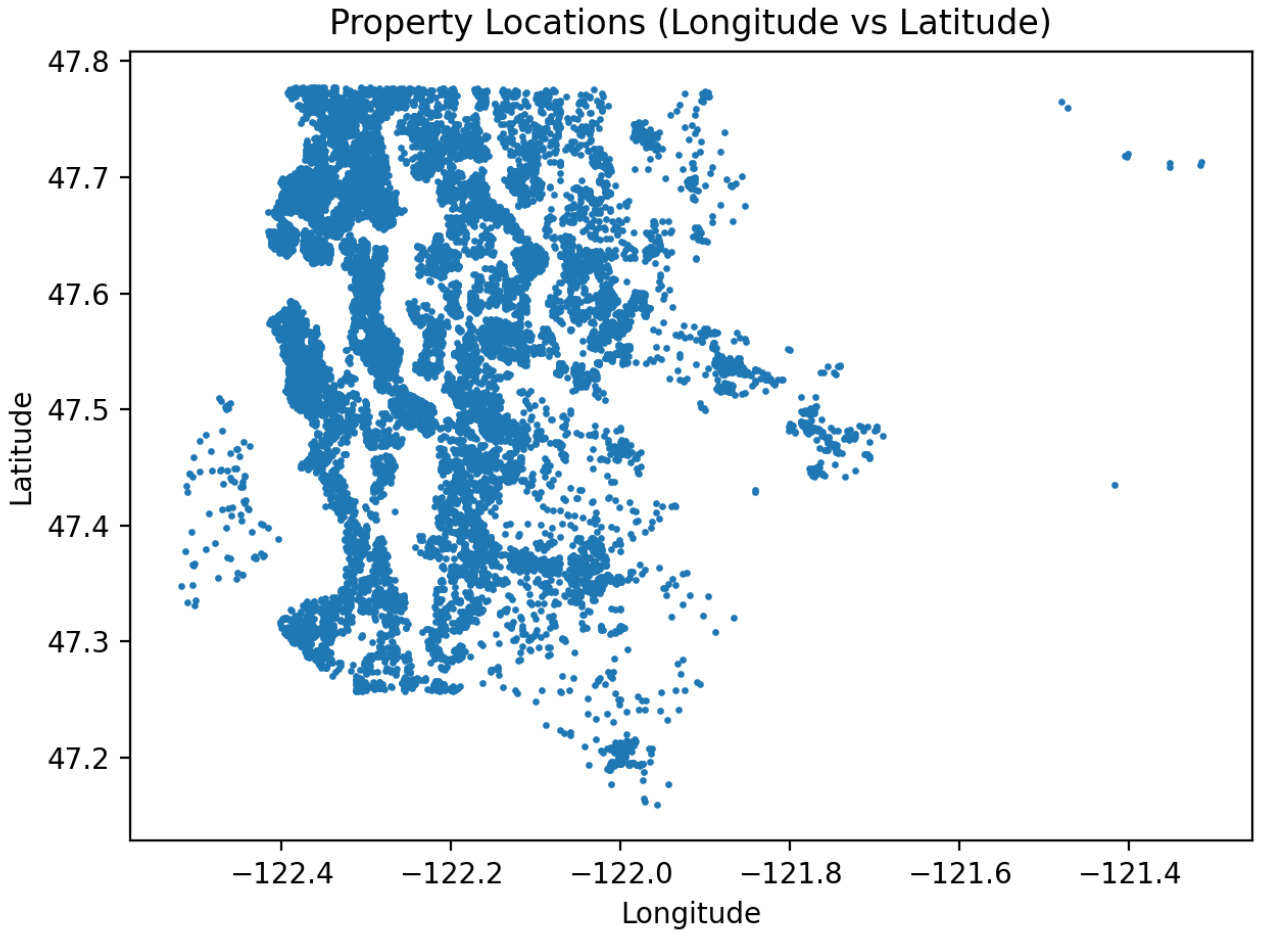
- Original prices are highly right-skewed
- Log transformation stabilizes distribution, Model trained on log(price + 1)





EDA: Spatial Distribution

- **Why location matters**
- Properties cluster in urban/subur regions
- Spatial patterns motivate visual context modeling



Satellite Image Exploration

What satellite images capture

- Green cover
- Road connectivity
- Urban density
- Waterfront proximity

Sample satellite images from training set

Financial & Visual Insights

Insights learned by the model

- High greenery → higher property value
- Better road connectivity → premium pricing
- Waterfront proximity → strong price uplift
- Dense concrete regions → relatively lower valuation

Modeling Strategy

Baseline Approach

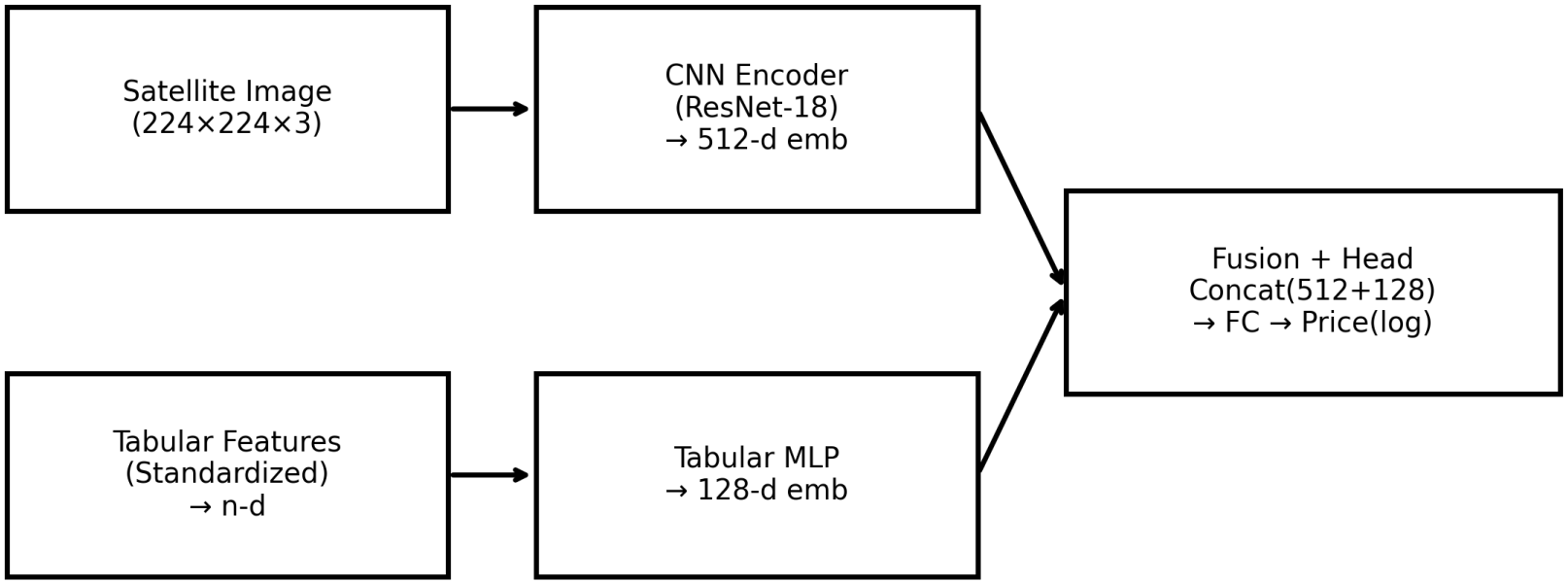
- Tabular-only regression model
- Strong but limited to structured features

Proposed Approach

- Multimodal learning:
  - CNN for images
  - MLP for tabular features
  - Feature fusion for prediction

Model Architecture

Multimodal Regression Architecture (CNN + Tabular Fusion)



Architecture Components

- Image Encoder: ResNet-18 (pretrained)
- Tabular Encoder: Multi-layer perceptron
- Fusion: Concatenation + Fully Connected layers
- Output: log-price prediction

Training Details

Training Setup

- Loss: Mean Squared Error (log-space)
- Optimizer: AdamW
- Hardware: GPU acceleration
- Validation: 80–20 train–validation split

Why log-price?

- Handles skewed price distribution
- Improves training stability

Performance Comparison
Evaluation Metric- RMSE (log-price)

Model	Input	RMSE (log)
Tabular Baseline	Structured only	~0.33
Multimodal Model	Tabular + Images	**0.289**

Result Interpretation

What does RMSE = 0.289 mean?

- Average prediction error ≈ **30–35%**
- Significant improvement over tabular-only model
- Confirms value of satellite imagery

Key takeaway

- Visual environmental context meaningfully improves property valuation.

Conclusion

Conclusions

- Multimodal learning outperforms traditional models
- Satellite imagery captures valuable neighborhood signals
- CNN + tabular fusion is effective for real estate analytics

Future Work

- Higher-resolution imagery
- Temporal data (price trends)
- Grad-CAM based explainability

Thankyou