Predicting House Prices with Image Detection

Jonathan Wang
Grinnell College
Grinnell, IA, USA
wangjona@grinnell.edu

Ryan Guo
Williams College
Williamstown, MA, USA
ryg2@williams.edu

Kaitlyn Chen
Harvey Mudd College
Claremont, CA, USA
kaichen@hmc.edu

***Abstract*—For our project, we will develop a model for multi-step time series prediction of the sale price of houses based on image detection. We are using multiple datasets to train our model. two of our datasets have data on the prices of the houses, while two do not. The data encompasses over 20,000 images of home exteriors and interiors, and attribute data such as city, number of bedrooms, number of bathrooms, and square footage (in feet).**

I. Introduction

We developed a model for multi-step time series prediction of the price of houses based on image detection and numerical information. This is useful for people to want alternative sources other than Zillow and Redfin. This project helps inform decision making on purchasing houses, home renovation, and real estate.

II. Previous Work

A. Vision-based housing price estimation using interior, exterior & satellite images

This study addresses the limitations of traditional house predicting methods that rely on text data by incorporating images [2]. These images are of the house's interior, exterior, and satellite images of a house to predict the price of it using convolutional neural networks (CNNs). It also uses the number of bedrooms, number of bathrooms, number of square feet, age, and zip code. The CNNs can detect features of the house. The images were classified based on their luxury. For the interior rooms, each room was evaluated separately. The luxury levels were crowd sourced by a random group of people who selected rooms of similar luxury levels. They then trained VGG16 for classification of images, which had 93% validation accuracy. The luxury levels and meta data were concatenated to form as the features of each data point. They used Gradient boosting regression model.

B. House Price Prediction via Computer Vision

This project predicts house prices from exterior frontal images of houses using CNN and various preprocessing techniques. The dataset they used contains over 20,000 images and numerical data for single-family homes in Southern California. The model, trained on this dataset, can predict house prices with an average error of 4.3%. This experiment has limitations in terms of computational power, the need for more images (both interior and exterior), and the tuning of the CNN model [3].

III. Dataset

To retrieve the appropriate data to produce insightful metrics, we searched for datasets consisting specifically of house images with corresponding quantitative and categorical data labeling. We found four datasets for our model. The first dataset is a Kaggle dataset called *House Prices and Images - SoCal*, which has $\sim$15.5k images of the exterior of the house and a CSV that contains the columns: city, number of bedrooms, number of bathrooms, number of square feet, and price [2]. Another dataset is from a project on Github, titled *Houses Dataset* [1]. It has images of the bedroom, bathroom, kitchen and house front, as well as information for the number of bedrooms, bathrooms, area, zip code and price. The third dataset we use is *House Rooms Images Dataset*, which has bathroom, dining room, living room, bedroom, and kitchen images, with no information on the house prices [4]. The fourth dataset is called *House Rooms & Streets Image Dataset* and has images of the bathroom, dining room, living room, bedroom, and kitchen image (like the previous dataset), as well as street images like of apartments, churches, garages, houses, industrials, office buildings, and roofs [3]. We prepare this data by combining similar rows and columns and combining the image datasets, as well as getting rid of columns that wouldn't overlap across datasets and rows without information. We ensure all listings in the CSVs have corresponding images and remove any unmatched records.

IV. Method

We leverage a combination of datasets: smaller datasets containing categorical and quantifiable data along with larger datasets centered around a mass collection of images. Our system implements a core strategy of two-phased learning, combining unsupervised learning on image-oriented datasets with supervised fine tuning on information-based data to maximize our utility of information:

Phase 1: Learning General features

Our model analyzes images from *Houses Rooms Image Dataset* and *House Rooms Image Dataset* to enable learning of universal patterns such as room layouts, architectural styles, and material quality. We integrate techniques such as contrastive learning to train the model to recognize similar and dissimilar houses and build a visual intuition of house images.

Phase 2: Fine Tuning for Price Prediction

From Phase 1, our pretrained model is adapted to predict sale prices using the datasets containing quantifiable and categorical information (*SoCal* and *Houses Dataset*). The model's early layers at this step are frozen to preserve learning features, isolating the final layers to map these features to dollar values. By initializing pretrained features, we ensure the model requires fewer priced examples to maintain good accuracy.

Baseline Model

To further integrate the image-based data, the system employs multi-task learning to simultaneously predict prices alongside room types (e.g bedrooms/kitchens). The implementation of the baseline model goes as follows: Initialize a Random Forest Regressor model. Train on combined features, extracting image features if available. Otherwise, storing into a variable containing found tabular features. As such, we designed the model to specifically learn features useful for the prediction of sale prices and room types, improving on generalization. By incorporating *pseudo-labeling*, we train and initial model on priced data, use this to predict prices on housing images, then retrain the model using high-confidence predictions as additional training data.

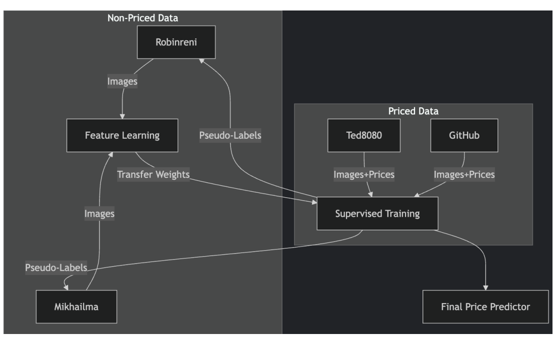


Diagram of baseline model architecture

Deep Learning (CNN) Model

The training process flow of the CNN deep learning model begins with the priced data being split into train, validation, and test sets and the preprocessing of the nonpriced data into image augmentations. Augmentations were applied to image-based data such as random crops, color jitter, and horizontal flips. Contrastive learning was then applied to collect information on general visual features and robustness to lighting and viewpoint variations. Supervised fine tuning on the priced data followed by a semi-supervised refinement was also executed before evaluation.

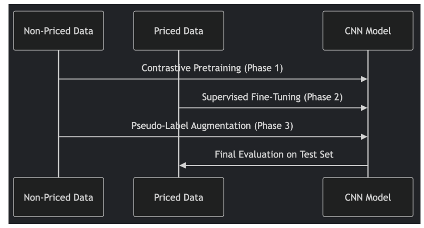
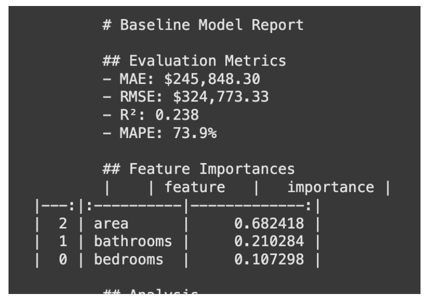


Diagram of CNN model Flow Chart

V. Evaluation

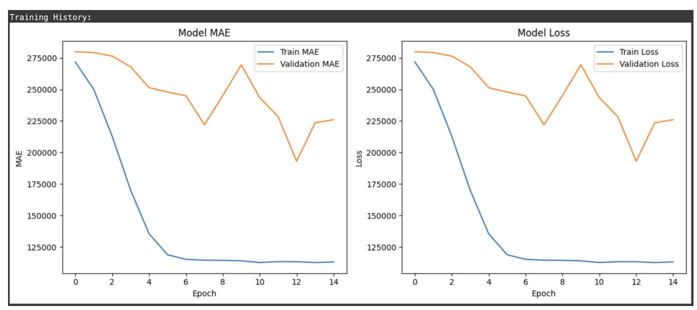
The output metrics of our baseline model displayed MAE 245,848 and RMSE 324,773 error prices. our R^2 value was 0.238 and the MAPE value was 73.9. The model also determined the area, bathrooms, and bedrooms as the most important features.


```
# Baseline Model Report

## Evaluation Metrics
- MAE: $245,848.30
- RMSE: $324,773.33
- R²: 0.238
- MAPE: 73.9%

## Feature Importances
|    | feature   |   importance |
|---:|:----------|-------------:|
|  2 | area      |     0.682418 |
|  1 | bathrooms |     0.210284 |
|  0 | bedrooms  |     0.107298 |
```
Baseline model evaluation

After evaluating the CNN model across 20 epochs, we obtained an MAE price of 193045 and an RMSE value of 229320. We also display training history charts based on the MAE and loss metrics:



History charts of MAE and Loss metrics

VI. Results

Overall, the CNN model outperformed the baseline Random Forest model in terms of MAE and loss metrics, indicating that the deep learning approach provided a modest improvement in predictive accuracy. The similarity observed between the training loss and MAE curves is likely because the Huber loss function, which interpolates between MAE and MSE, produces values that can be numerically close to MAE, especially when prediction errors are relatively large. This causes the loss and MAE metrics to track similarly during training.

VII. Discussion

While our current models demonstrate some insightful results in a methodology of predicting house prices using both tabular and imaging data, the overall performance indicates room for significant improvement before reaching a level suitable for practical application. The relatively modest gains of the CNN model over the baseline suggest that the integration of image data has potential but requires further refinement.

VIII. Summary

Our project aimed to investigate the prediction of house prices by leveraging both tabular data and imaging through machine learning combined with deep learning approaches. We compiled four diverse datasets containing several related house images and associated quantitative and categorical features, preprocessing each of them to feed into our models. Our methodology involved a two-phase learning system using unsupervised and supervised fine-tuning. We developed a baseline Random Forest regression and a CNN model that incorporated image augmentations and semi-supervised learning. Evalutation displayed that the CNN modestly improved on MAE and RMSE values, highlighting an impact of including image data. Further refinements to adjust our application to this project idea could lead to a more robust prediction.

References

[1] H. Ahmed E. and Moustafa M. (2016). House Price Estimation from Visual and Textual Features. In Proceedings of the 8th International Joint Conference on Computational Intelligence (IJCCI 2016)ISBN 978-989-758-201-1, pages 62-68. DOI: 10.5220/0006040700620068

[2] A. Nouriani, L. Lemke (2022). Vision-based housing price estimation using interior, exterior & satellite images, Intelligent Systems with Applications, Volume 14, 200081, ISSN 2667-3053, https://doi.org/10.1016/j.iswa.2022.200081.

[3] tncy67. (2022). House-Price-Prediction-via-Computer-Vision [Computer software]. GitHub. https://github.com/tncy67/House-Price-Prediction-via-Computer-Vision

[4] Mazurov, M. (2022). House Rooms & Streets Image Dataset [Data set]. Kaggle. https://www.kaggle.com/datasets/mikhailma/house-rooms-streets-image-dataset

[5] Reni, R. (2024). House Rooms Image Dataset [Data set]. Kaggle. https://www.kaggle.com/datasets/robinreni/house-rooms-image-dataset