Predicting Prices of Taiwanese Commercial Real Estate

Sharlene Chen, Shih-Ya Chang, Vivian Lee, Jenny Wei

Introduction

Issue

- Common commercial real estate price predictions rely merely on traditional factors such as internal conditions of the properties and subjective opinions from realtors. As a result, price information is usually asymmetric and predictions usually lack objectivity. Apart from traditional factors, external features such as parks, restaurants, and road feature could impact the price. Therefore, we aim to include objective features and form a model that can predicts more accurate commercial property prices.

Hypothesis

- Both internal and external features play essential roles in the real estate market.

Expected result

- The deep learning model can predict commercial real estate price based on both internal and external features.
- The combination of both features can predict commercial real estate price more precisely.

Dataset

- We first retrieved Real Estate Transaction data in Taipei and New Taipei city [1]. This dataset has various features such as building age, number of floors, building material, number of rooms, etc.
- We utilized Heremap and Google APIs to add corresponding longitude, latitude, and nearby features (e.g. universities, stores, travel agencies) to our data.
- In addition, we used Mapbox API to collect map images around each data (Size: 256 x 256).

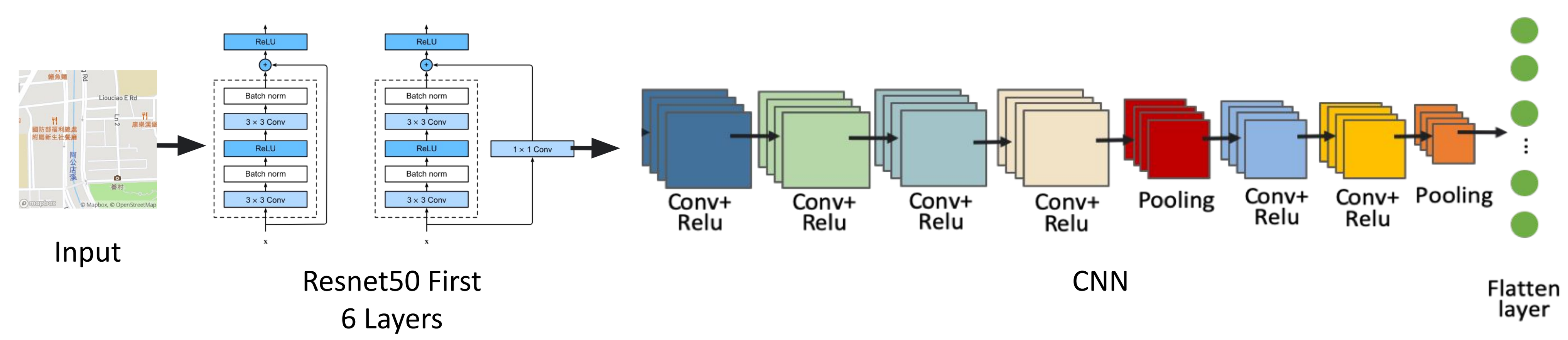
Methodology and Results

Data preprocessing

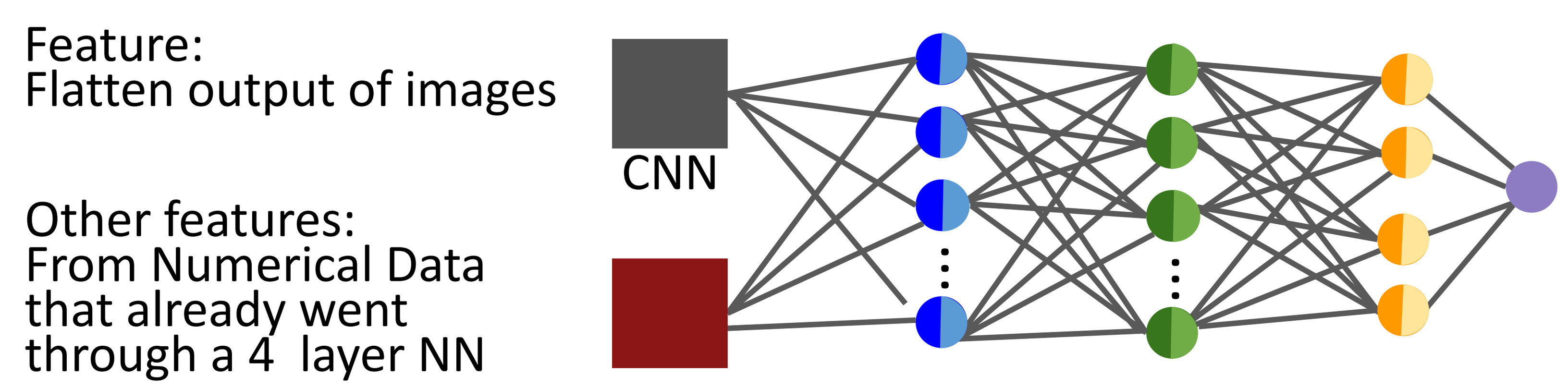
- Numerical Dataset:
  1. Convert string to numerical value(int/float/double..)
  2. Deal with missing data
  3. One-Hot Encoding
  4. Standardization
- Image Dataset:
  1. Typecast each RGB image (mapbox images) to a tensor (3*256*256)

Algorithm and Training

- Resnet50 + CNN (Convolutional Neural Network):

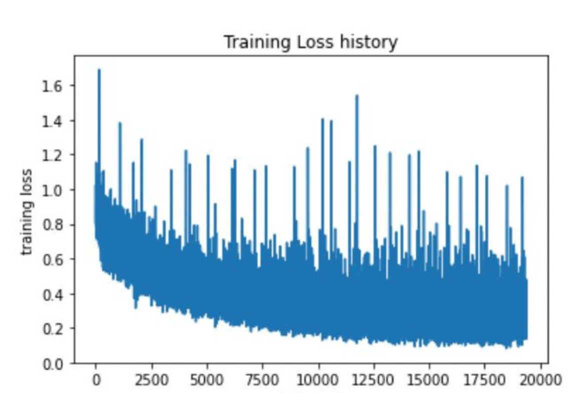


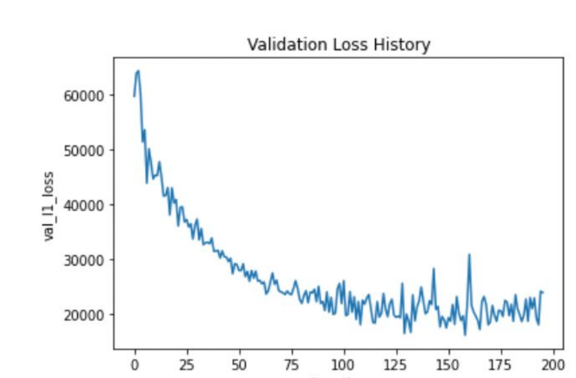
- DNN (Deep Neural Network)



Testing and Recommendation System

Best Result:
+- $13,000/sqm
($39,000/ping)





Discussion and Conclusion

- Numeric model has smaller L1 Loss but trains slower.
- Concatenated model trains faster but has larger L1 Loss.
- Both numeric and concatenated models trained L1 Losses smaller than one standard deviation.
- Images from street map, rather than satellite map, may cause larger L1 Loss.
- Data variability such as price differences in different cities may cause larger L1 Loss.

Future Work

- Improve concatenated models to reduce loss
- Change images to satellite map
- Reduce data variability by limiting region included
- Train models that can predict commercial prices in different cities/countries
- Explore the impact of natural/regional environment on commercial real estate prices by controlling inner or outer variables (features)

References

[1]Understanding and Visualizing Resnet
[2]Using Convolutional Neural Networks to detect features in satellite images
[3] Real estate transaction data
[4]Real Estate pricing with Machine Learning & non-traditional data sources

stanCode
標 準 程 式 教 育 機 構