MIE324

Project: Uncle Penny Bags

Final Report
1678 words

December 2nd, 2018

D. Aracely Escoboza
Oliver Liang

Goal and Motivation

The motivation behind our project comes from the fact that nowadays the process of buying/selling a house can be very costly and time consuming for people. In addition, a lot of houses are not listed in real-estate websites so the necessity of getting a professionals' opinion becomes bigger.

Our main goal is to provide possible home buyers and realtors in Canada with an algorithm that will facilitate the process of analyzing a property to obtain its value. The algorithm will work such that when given images of a property and its location, it will output the estimate of the property value.

Overall Project Structure

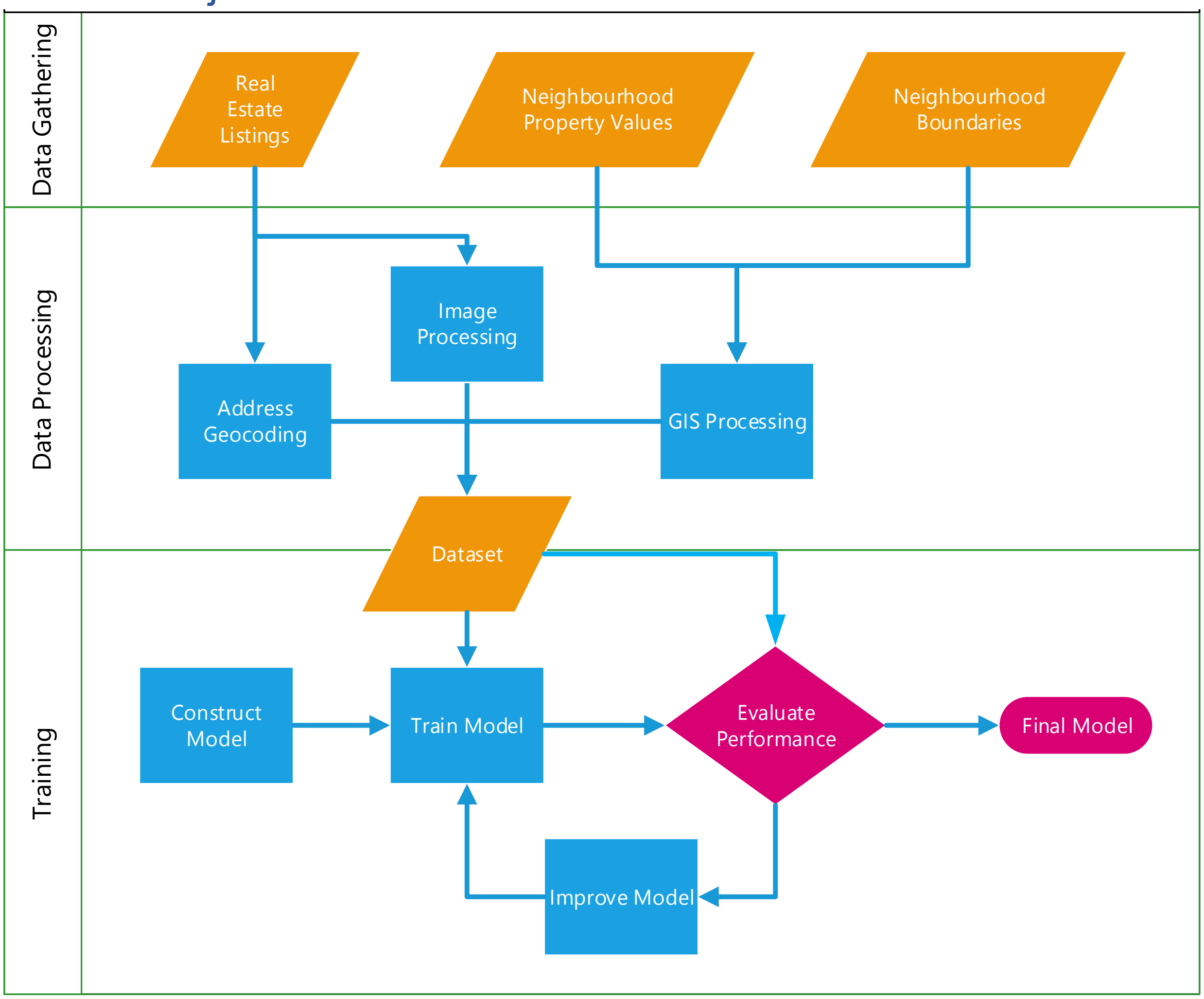

Figure 1 | Project structure and workflow

Data Gathering and Processing

For this project, our data came from 2 main sources: zoocasa.com [1], and the National Household Survey (NHS) from the 2011 census [2].

Zoocasa.com was used to provide the real estate listings and images for training the predictor.

A scraping program was built using python, selenium web-driver and Google Chrome, in order to extract the address, price, and images associated with each listing. The dataset for this project was taken on Nov 11, 2018, from the first 800 pages of the houses category under zoocasa.com/houses. The resulting 10,000 initial listings was then filtered down to about 9000, by only selecting listings with valid prices, and 2 or more images.

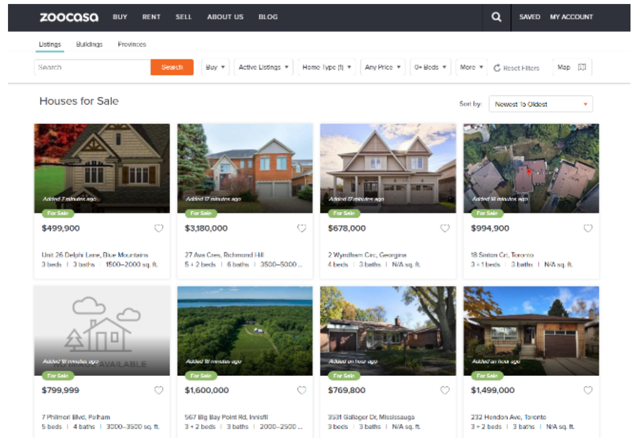

Figure 2 | zoocasa.com

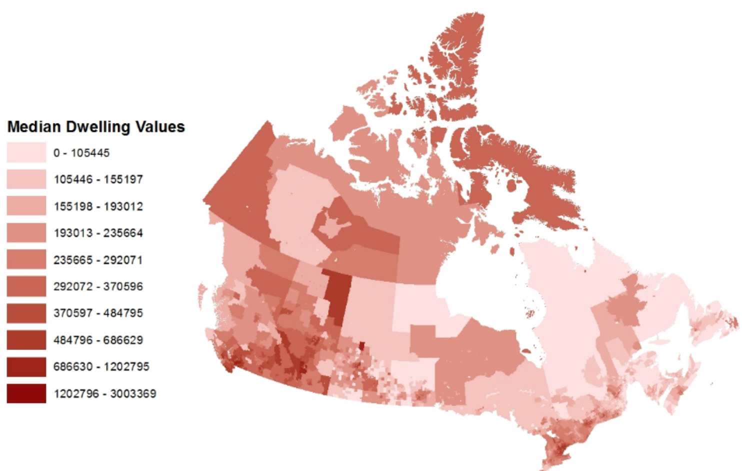

Figure 3 | Map of Canadian Median Dwelling Values from 2011 Census

As the real estate market varies significantly from place to place, we decided to anchor the predictions by only predicting prices as linear multiple of neighbourhood property values. The source of neighbourhood property values we used was the mean and median dwelling values from the NHS. The dwelling value data is provided in tabular form and is related to census areas. ArcGIS was used to relate the tabular NHS data with 2011 census boundaries [3]. This resulted in a vector database (Figure 3) that was imported into python using Cartopy. The neighbourhood property values was then determined for each real estate listing by first converting the address to longitude and latitude using the Google Geo-coding API, then looking up the coordinates in our vector database.

And lastly, in order to argument the data and speed up the training process, the images from all listing was randomly resized, cropped, horizontally flipped, then saved, 4 times into colour 224x224 jpgs, and center cropped and saved once for validation.

The final dataset was saved in a JSON file as a list of real estate listings with the following format:

```
[

      {
     'images' : [
         'image_name1.jpg',
         'image_name2.jpg',
           .
           .
      ],
      'median': neighbourhood median value (CAD),
      'mean': neighbourhood mean value (CAD),
      'price': property list price (CAD)
      },
      .
      .
      .
]
```

Machine Learning Model

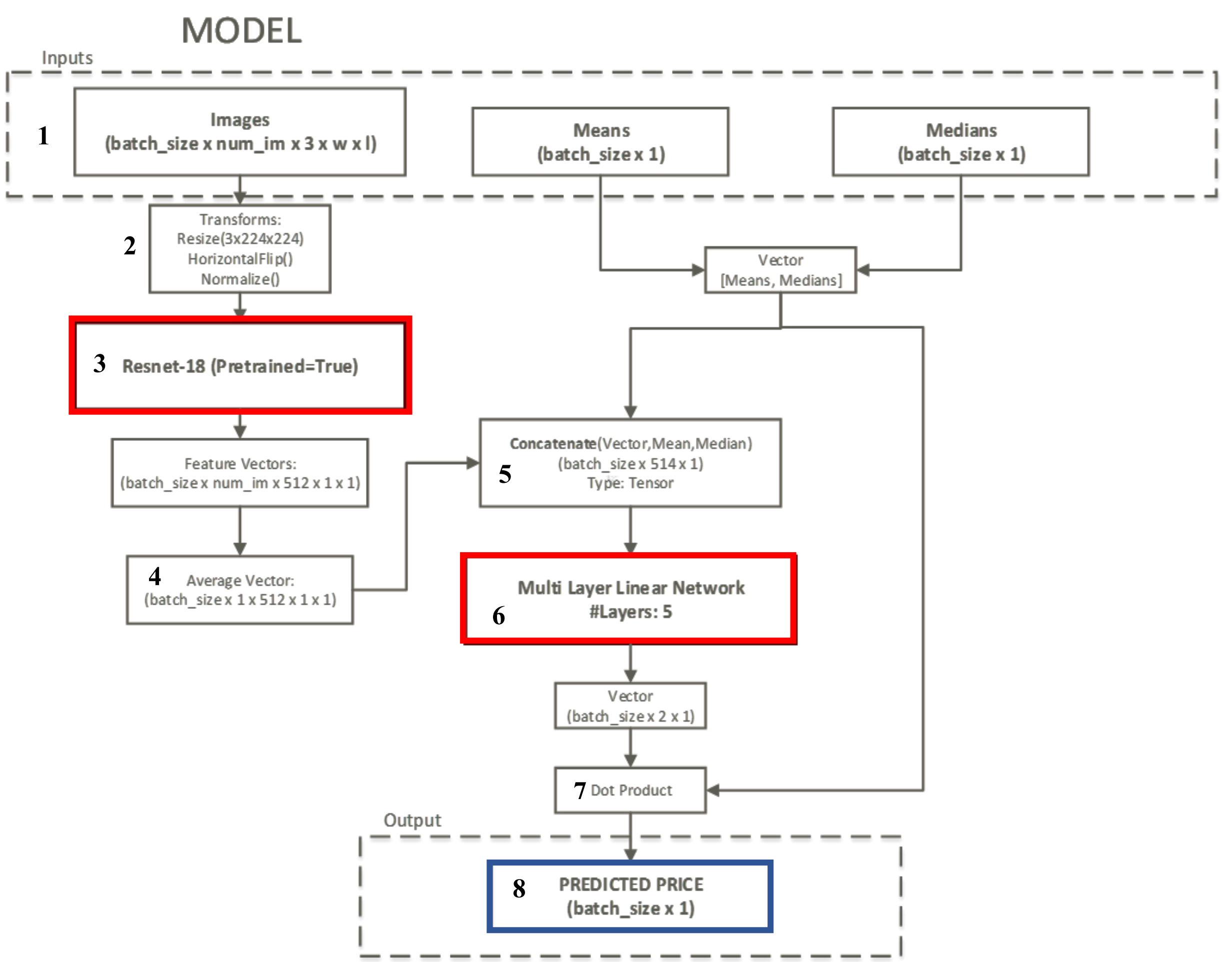

Figure 4 | Machine Learning Model

Model Notes

- The inputs of the model consisted of the images of the house, the mean and the median neighbourhood values (1). The images were transformed (resized and flipped) and saved before running the training loop (data augmentation method used) (2).
- After, the normalized images got feed into the pre-trained Resnet-18 (3) and this outputted a feature vector for each image. Since each house had multiple images and thus multiple feature maps, we decided to take the average feature map (4) for each house in order to continue to the next steps.
- When we had the 512x1 feature vector for each house we concatenate it with a scaled version of the mean and median values of each house's neighbourhood to obtain a 514x1 vector that will be fed to the linear network (5).
- This vector went through a 5-layer linear network with Relu activation functions in between until arriving to a 2-dimensional vector that went through a final activation function: Softplus (6). The vector was finally used to take the dot product (7) with the mean and median values to arrive to the final predicted price (8).

Training Process

Accuracy Criteria:

To consider the prediction to be acceptable, in comparison with the actual value of the house in the real-estate database (label), we decided to judge the prediction on a range from the actual value, where if the prediction it's within the deviation boundaries it will be considered as an acceptable (right) answer.

After consideration we decided to set a criteria of ± 20% of the true value. We see a prediction within this range as an acceptable answer because in the process of buying/selling a house the real selling price of a house can be drastically different form the listing price [4]. The fluctuation within the house-market is also a reason why predicting the actual value/worth of a house can become a very complex problem.

Initial Hyper Parameters used

- Learning rate: 0.001
- Batch size = 20
- Epochs = 50
- Training data portion of the total data = 80%
- Resnet-18:
  - We removed the last linear layer
  - Pre-trained
  - Frozen during first half of training
  - Unfrozen for the last half
- Multi Linear Layers:
  - 5 linear layers
  - 4 Relu activation functions
  - Last activation function: Softplus (smooth representation of Relu)
- Optimizer: Adam
- Loss function: SmoothL1Loss
  - The predicted value was divided by the truth and compared against 1
- Training / Validation Split: 80% - 20%

Reasoning

Why Resnet-18?

- 1st place in the ILSVRC 2015 classification competition.
- Resnet was able to solve the problem of degradation that happens when a network gets deeper, solved by using residual blocks [5].
- It's the only 2 Resnets that fit in the 1 GPU's memory
- Use a pre-trained network to do transfer learning, and reduce data requirements

Why use Mean and Median as an input to linear layers and then again for the final computation of the prediction?

- Using the mean and median values of the neighbourhood properties database gives our model an insight of the correlation that there is between the location and the worth of a property.
- Scaling the mean and median values before concatenating them with the feature image vectors helps the neural network perform better because the image vector contains small values.
- Using the mean and the median to calculate the predicted price with the output of the linear neural network helps us be able to maintain the output of the network in small numbers (since linear layers work best with smaller value predictions). Also, since the network already has an idea of the mean and median values from the input, it gives it an insight of the correlation that can be between the output and the mean and median.

Why scale the predicted value?

- As property value increases, so does the variability in its perceived value between sellers and buyers.

Why freeze the Resnet at the start?

- So, the linear layer can start by learning from pre-trained ImageNet 'embeddings' in the resnet's output. This should also allow for more consistent training results

Initial training results

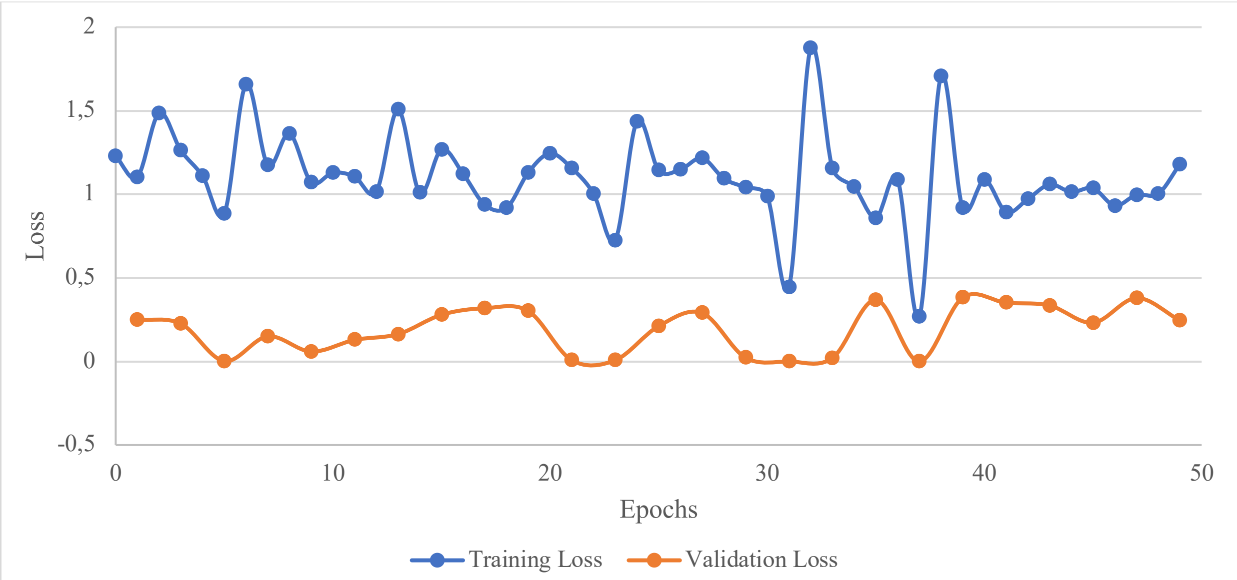

Figure 5 | Initial Training Vs Validation Loss Graph

The initial model suffered from heavy instability, under fitting, and showed no improvement from start of the training process to the end. And it had a top validation accuracy of 42%

Final Hyper Parameters used

- Learning rate: 0.0001
- Batch size = 20
- Epochs = 100
- Resnet-34
  - Frozen only for the first 15 epochs
- Multi Linear Layers:
  - 1 Hidden Linear Layer
  - Tanh activation function
  - Soltplus on the output layer
- Only computing the price as a multiple of the neighbourhood mean, instead of linear combination of mean and median.
- Loss function: SmoothL1Loss
  - The predicted value and truth were divided by the neighbourhood mean then compared

Reasons for change

Why switching form Resnet-18 to Resnet-34?

- Resnet-18 was initially used as it was the only Resnet that could fit into GPU memory. But after optimizing the training loop, enough memory was freed up that resnet-34 could be trained at the same batch size, and it had better accuracy.

Why reduce the size of linear layers?

- After experimenting with different linear layer sizes, it was found that more than 1 hidden layer didn't improve the model performance.

Why remove neighbourhood median from direct model output?

- Removing it didn't impact the accuracy, and allowed for faster convergence

Why changing loss function input scaling?

- It seemed to dramatically improve the training speed and resulting accuracy.

Final result:

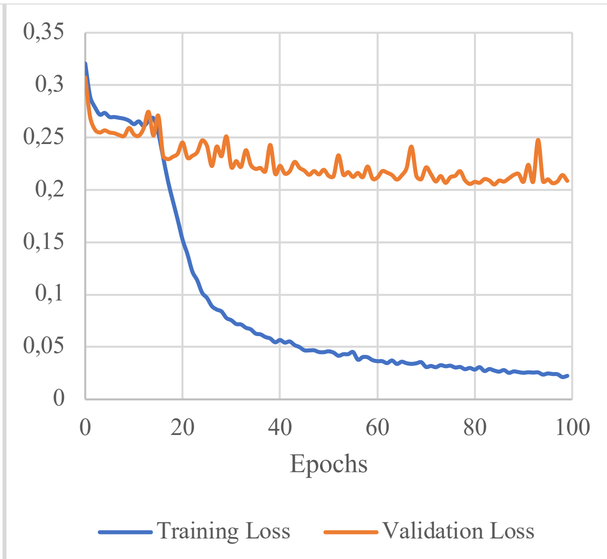

Figure 6 | Final Training vs Validation Accuracy

Figure 7 | Final Training Vs Validation Loss

The best validation accuracy of our final model was 49.2%. Although this was an improvement from the initial model, the problem of under fitting was changed to over fitting. This is likely due the model remembering the combination of images, as there are only around 9000 training examples.

Ethical and Legal Issues

- Too much trust from people to this algorithm may alter the housing market in unpredictable ways. An analogous situation is the current stock market. As much of the prediction and decisions are now made by black box algorithms, random flash crashes are happening more frequently [6]. The fact that the housing market is much less liquid than the stock market, would reduce this issue, but not eliminate it.
- If the population see the output of an algorithm like this as the ultimate truth, we can run into problems of people trying to change their home to satisfy the needs of an algorithm, instead of actual buyers.
- The ownership of such an algorithm would be under question, as a fully trained model would likely incorporate the listing from several real estate websites. And if the predictions are used for profit, some of the sites might claim the model as their intellectual property.

Key Learnings

- The transformations made on the images are computed faster when performed inside of the model then when being performed before loading.
- The importance of proper identifying which operations should be computed with a GPU/CPU in order to save computation time.
- Neural Networks don't work well with big numbers for either input or output.
- Half precision floats are not a good idea for regression models.
- GPU memory is valuable.

References

[1] Zoocasa, "Houses for sale in Canada - Search MLS," Zoocasa, [Online]. Available: https://www.zoocasa.com/houses. [Accessed 14 November 2018].

[2] Statistics Canada, "National Household Survey (NHS) Profile, 2011," 8 May 2013. [Online]. Available: https://www12.statcan.gc.ca/nhs-enm/2011/dp-pd/prof/details/download-telecharger/comprehensive/comp-ivt-xml-nhs-enm.cfm?Lang=E.

[3] Statistics Canada, "Statistics Canada, Boundary Files," 27 July 2012. [Online]. Available: http://geo.scholarsportal.info/#r/details/_uri@=3789411615$DLI_2011_Census_DBF_Eng_Nat_csd. [Accessed 12 November 2018].

[4] Google, "house sold above asking site:blogto.com," Google, [Online]. Available: https://www.google.ca/search?q=house+sold+above+asking+site%3Ablogto.com. [Accessed 1 December 2018].

[5] P. Jay, "Understanding and Implementing Architectures of ResNet and ResNeXt for state-of-the-art Image Classification: From Microsoft to Facebook [Part 1]," Medium, 7 Febuary 2018. [Online]. Available: https://medium.com/@14prakash/understanding-and-implementing-architectures-of-resnet-and-resnext-for-state-of-the-art-image-cf51669e1624. [Accessed 30 November 2018].

[6] AvanidharSubrahmanyam, "Algorithmic trading, the Flash Crash, and coordinated circuit breakers," *Borsa Istanbul Review,* vol. 13, no. 3, pp. 4-9, September 2013.