#1322: LEARNING URBAN PERCEPTION OF CITIES VIA STREET VIEW IMAGES

Lim Xin Qi, 1161303165

Supervisor: Dr. John See Su Yang, Moderator: Dr. Wong Lai Kuan

Abstract

The perception of humans, particularly of a city, is commonly coined "Urban Perception". In the past, the effort to quantify the urban perception of a city was tremendously tedious since the studies were mostly carried out manually. In this paper, a more efficient and data-driven way of urban perception study was carried using deep learning on city streetscapes to predict the perceptual scores of neighbourhoods in Kuala Lumpur, Malaysia.

Objectives

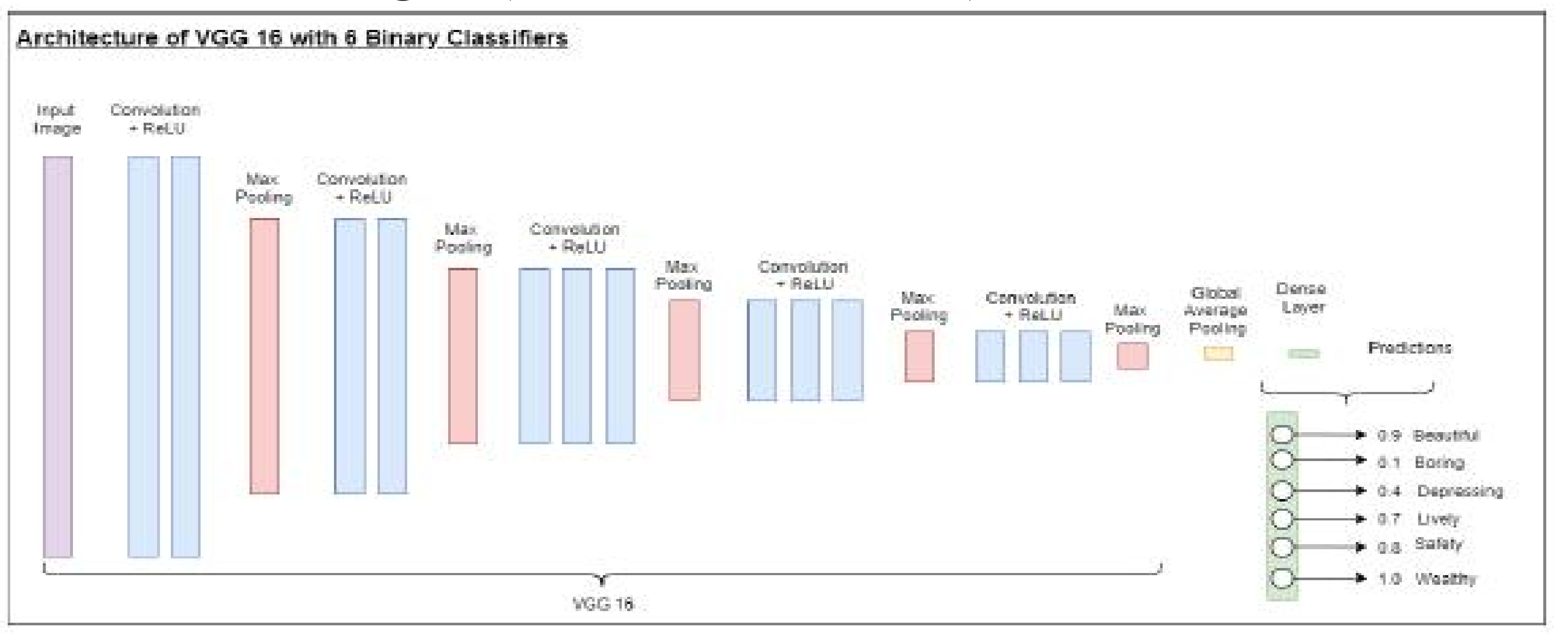
- To identify and train a suitable deep learning model for a multilabelled classification task to predict the urban perception of a location given the Google Street View (GSV) image of that certain location.
- To create an interactive neighbourhood-level visualisation on the predicted perceptual attributes.

Literature Review

- Salesses et al. (2013) 's work is one of the initial approaches on collecting a large scale of street view data for urban perception studies.
- Dubey et al. (2016) crowdsourced a new GSV dataset (Place Pulse 2.0). They implemented a Siamese-like CNN model to predict the winner in a pairwise comparison.
- Using Place Pulse 2.0,
  - Zhang et al. (2018) trained a SVM classifier to classify 6 binarised perceptual attributes on each image.

Methodology

- Dataset: Place Pulse 2.0 dataset (Dubey et al., 2016) - GSV images
- Model Architecture: VGG 16 architecture, pretrained on Places 365 weight (Zhou et al., 2017)



- Evaluation Metrics: mean average precision (mAP)
  - $AP = \sum_n (R_n - R_{n-1}) P_n$
  - Mean of AP
- Visualisation:
  - 6 Choropleth Maps
  - 1 Predominance Map
    - Predominant perceptual attribute of neighbourhoods
    - Strength of predominance:

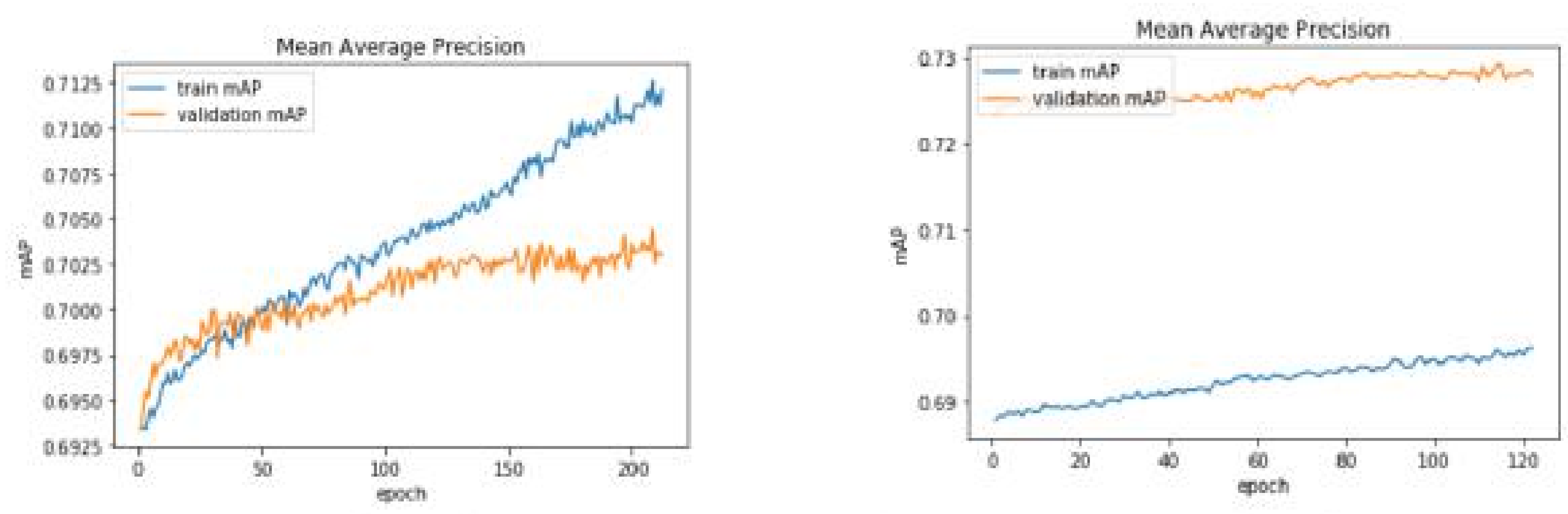
$$W_p = \frac{PerceptualScore_p}{\sum_{i=1}^{n} PerceptualScore_i}$$

Implementation, Findings & Result

- Training was done using different model settings while the architecture and weights used were all respectively VGG16 and Places365.
- The batch size specified was also constantly 16. The validation loss and validation mAP are both taken at the point of the lowest validation loss.

Model	Optimiser	Learning Rate	Data Augmentation	Validation Loss	Validation mAP
1	Adam	0.001	No	**0.84**	0.69
2	Adam	0.001	Yes	0.85	0.67
3	SGD	0.01	No	0.85	0.66
4	SGD	0.001	No	0.85	0.70
5	SGD	0.001	Yes	0.86	**0.72**

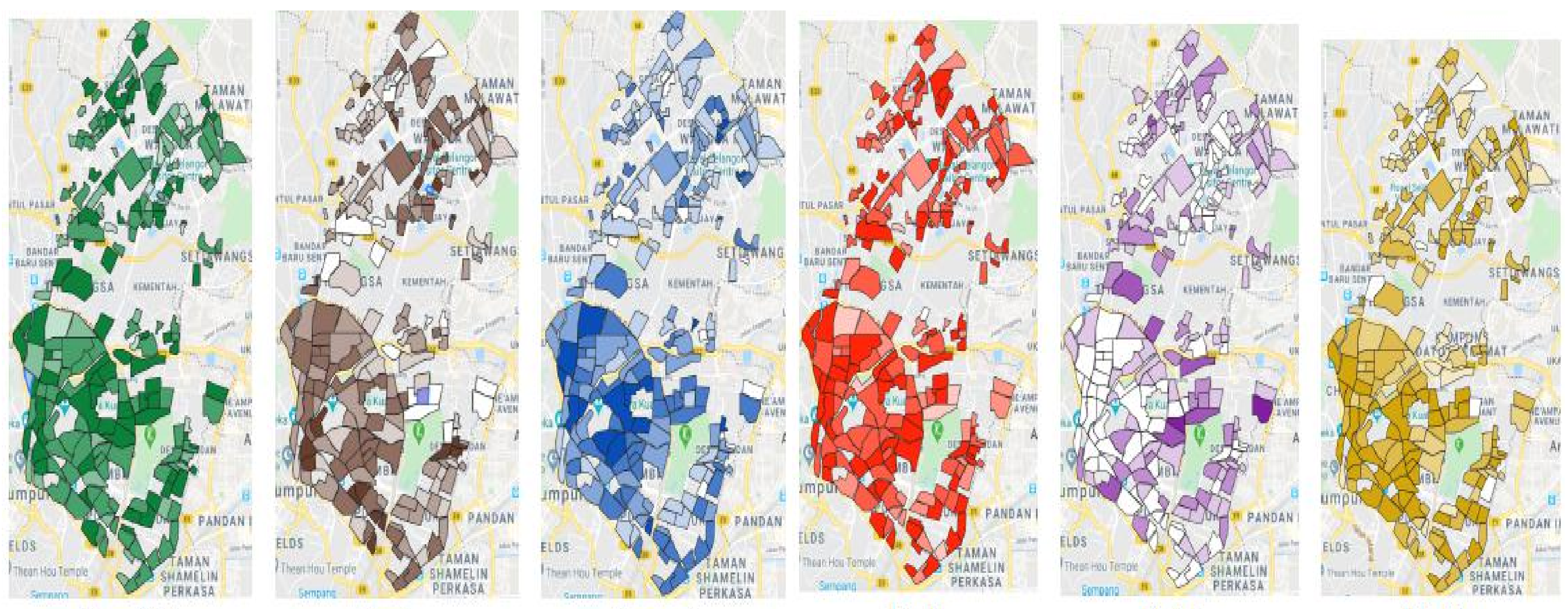
- Model 4 was chosen because it behaved more normally.



(a) Training mAP and Validation mAP of Model 4

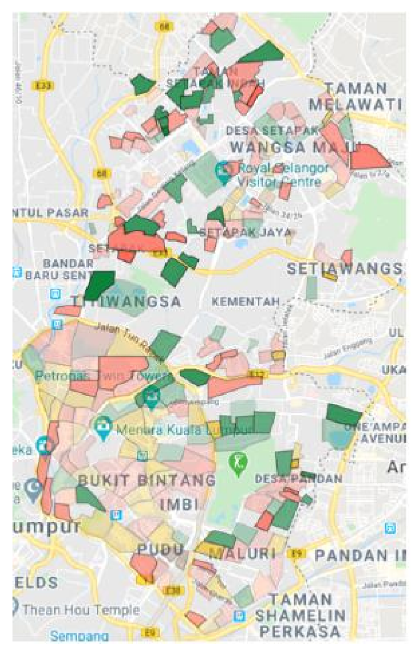
(b) Training mAP and Validation mAP of Model 5

- Choropleth Maps of 6 perceptual attributes



Beautiful Boring Depressing Lively Safety Wealthy

- Predominance Map
  - Color encoding: Predominant attribute
  - Transparency encoding: Strength of predominance



Conclusion

In conclusion, a suitable deep learning model which achieved a validation mAP of 0.70 was identified and trained for the multi-labelled classification task to predict the perceptual attributes for a location given the GSV image of that location. The predictions were visualised on a web-based platform which enables user interactions.