Systematic Street View Sampling for Accurate Urban Population Estimation

François Charih[1]* and Qinrui (Michelle) Si[2]

[1] Department of Systems and Computer Engineering
[2] Sprott School of Business
Carleton University
Ottawa, Canada

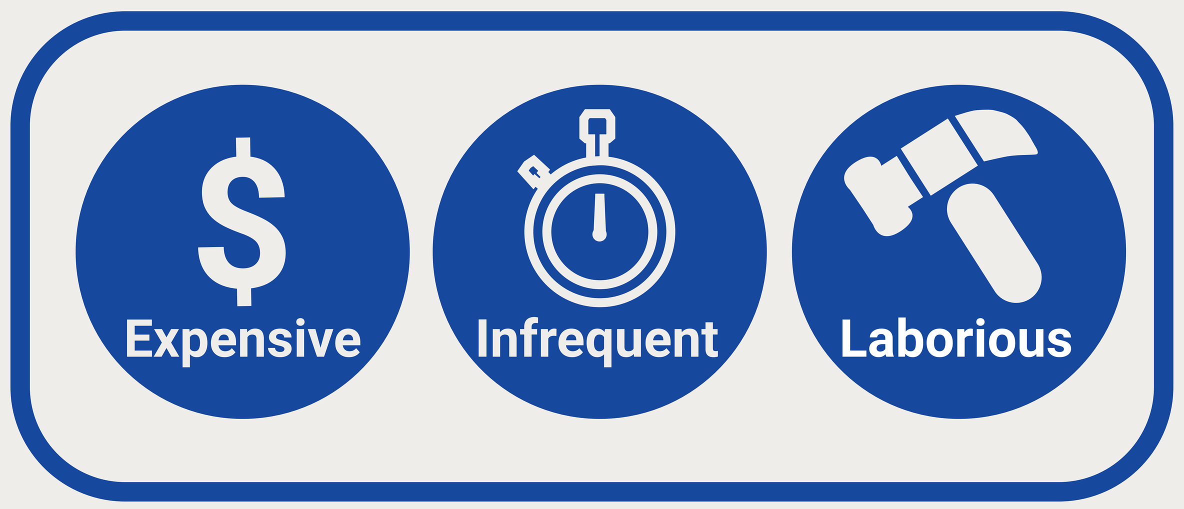
* Supervised by Prof. J. R. Green

Introduction

• Censuses provide rich information about a population and its demographic makeup.

• Policy makers and urban planners rely on population density estimates to optimize resource allocation and infrastructure development.

• Due to their cost, official censuses are only performed once every 5 years in Canada, and once every 10 years in the U.S.



Limitations of population estimation via census

• Good estimates of population count can be obtained by training deep learning models on satellite imagery [1].

• Street View imagery is expected to become increasingly abundant as self-driving cars reach the market. This provides a unique opportunity to study populations directly on the ground, at low cost and at fine-grained temporal resolutions.

• Groups have leveraged Google Street View (GSV) imagery to automate neighbourhood surveys and predict voting patterns [2, 3], but none have successfully used it to generate accurate population estimates, a logical first step in lowering census-related costs.

Reseach Objectives

Objective 1

Assemble a large, systematically sampled dataset of Street View imagery for multiple cities in the continental U.S.

Objective 2

Determine whether Street View imagery content can be leveraged to generate accurate population estimates in urban areas.

Objective 3

Investigate the generalizability of a model trained on U.S. city data to imagery from other countries.

Assembling a Dataset of Street View Imagery

• We **collected** a large quantity of imagery from 79 U.S. cities using Systematic Street View Sampling (S³) [4], an unbiased sampling algorithm with implementations that interact with Google's APIs.

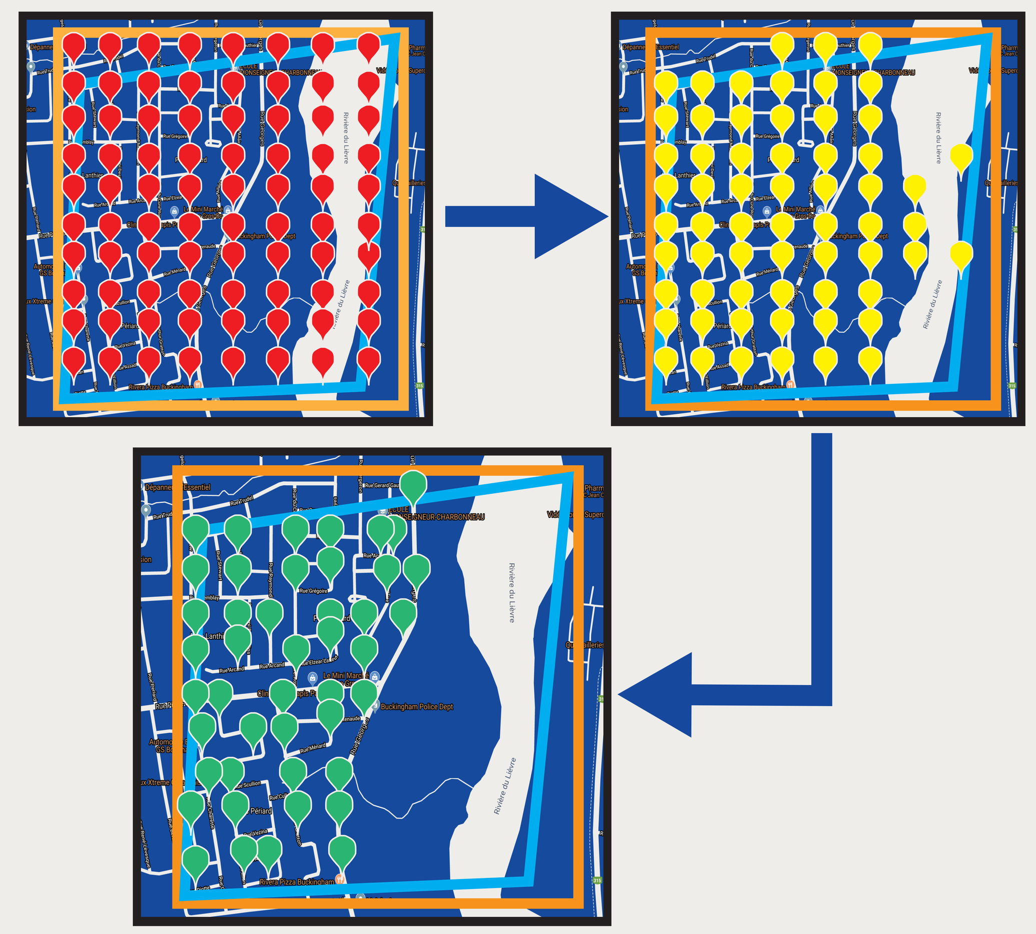


Figure 1. Systematic Street View Sampling (S³) algorithm. Candidates (red pins) are positioned along a grid defined by the region's bounding box (orange) and the sampling resolution. Coordinates outside the polygon (pale blue) or in water are removed. Remaining coordinates (yellow pins) are snapped to the nearest road. (green pins).

Sampling parameters

• Sampling resolution: **200 m**
• Number of headings per coordinate pair: **4**
• Pitch: **0°**
• Field of view: **90°**

Figure 2. Sample of 79 U.S. cities sampled for this experiment. The markers indicate cities for which Street View imagery was collected.

Deep Learning for Object Counting

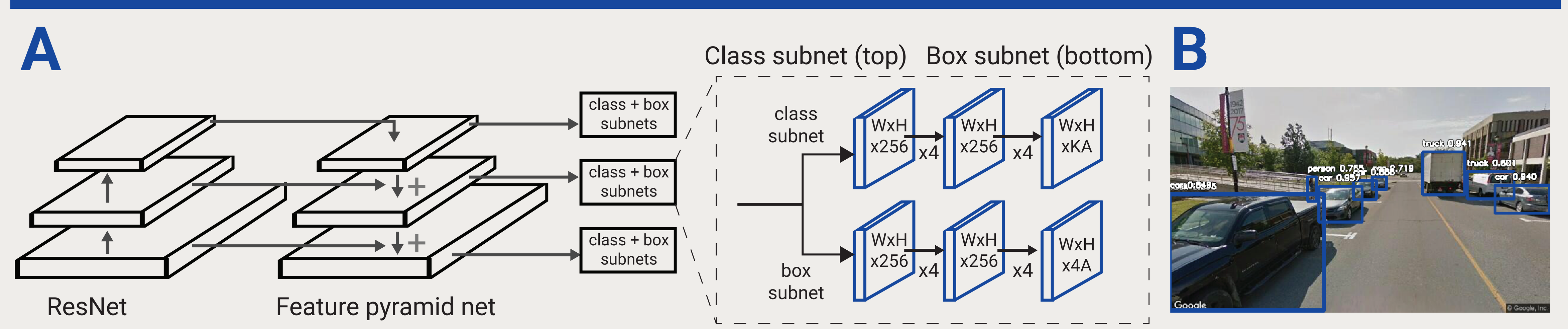


Figure 3. Fast object counting with the RetinaNet architecture. (A) Architecture of the RetinaNet convolutional neural network [5] used to count objects in Google Street View images. (B) Representative example of the application of the RetinaNet architecture on a Street View image taken on the Carleton University campus in Ottawa.

• **Deployed** a RetinaNet neural network architecture [5], implemented with Keras and pre-trained on the COCO dataset, on Compute Canada's infrastructure to make use of GPU resources.

• **Counted** 79 types of objects (people, cars, bicycles, trucks, trains, etc.) in over 425 000 images (resolution 640x380) from 79 U.S. cities.

• **Estimated** the accuracy of the object counts by crowdsourcing object counting tasks on Amazon's Mechanical Turk to estimate the accuracy of the RetinaNet network.

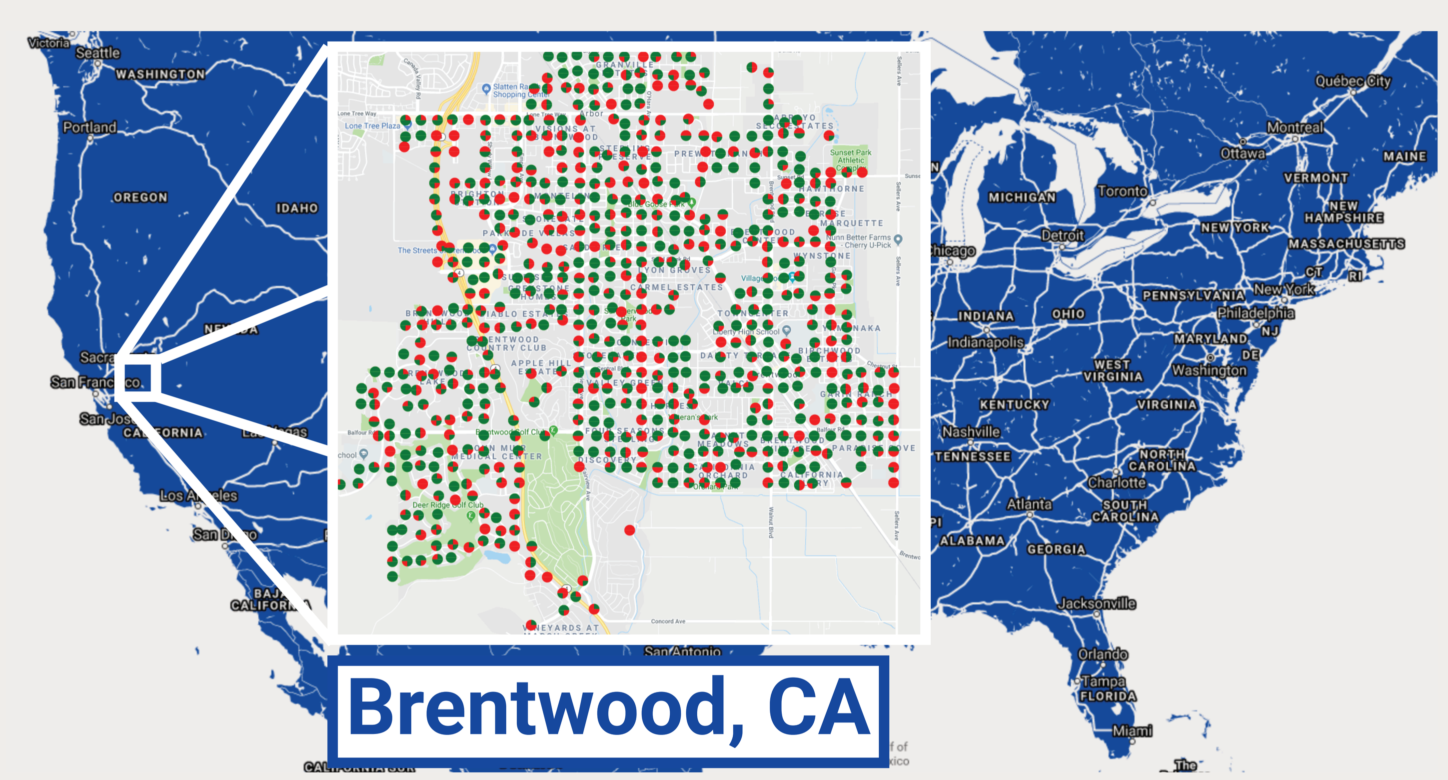


Figure 4. Street View imagery content from Brentwood, California. Circles correspond to coordinate pairs where imagery was obtained. The quadrants, corresponding to the four compementary headings of the images indicate whether objects were detected (green) or not (red).

Predictive Modelling

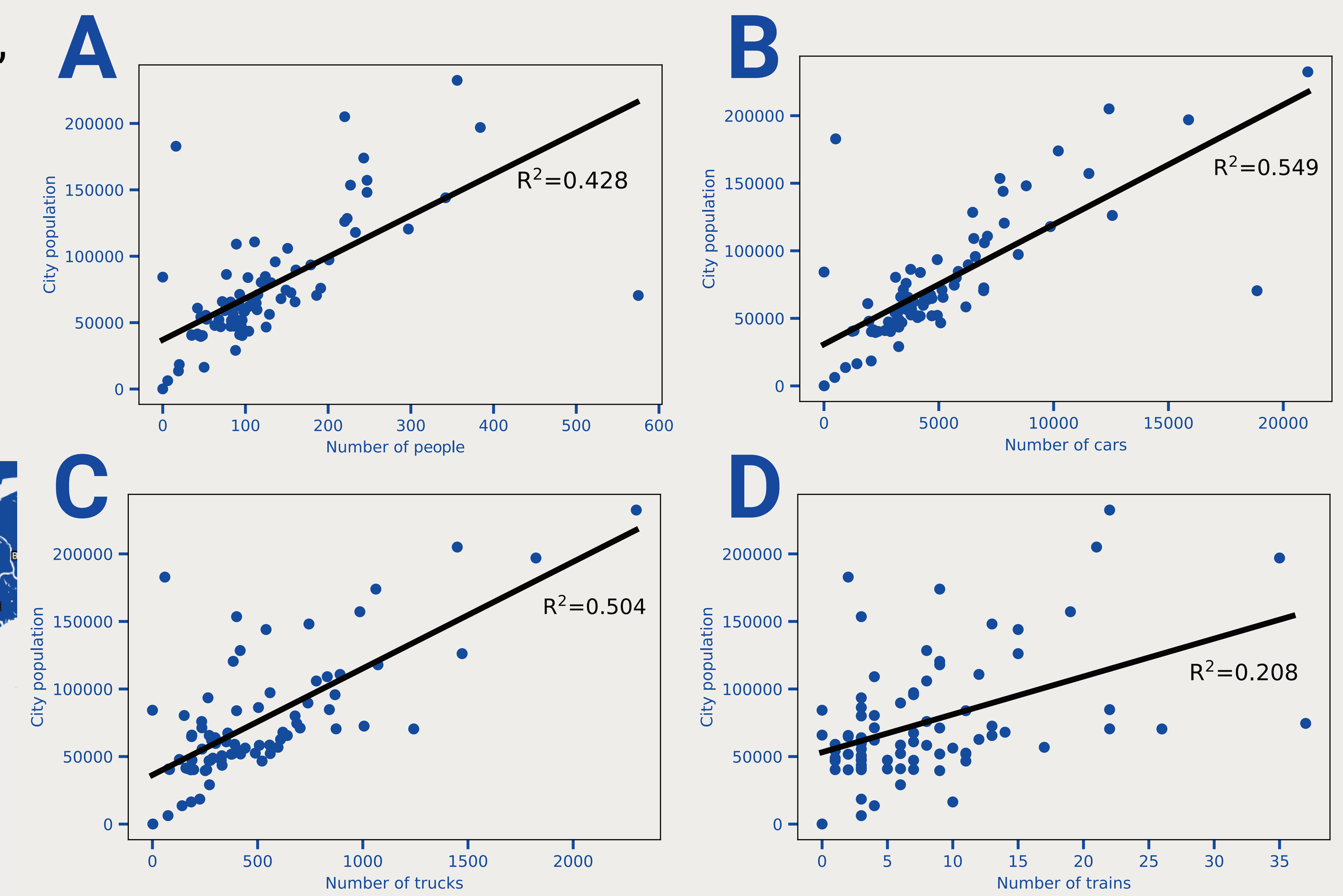


Figure 5. Correlation between the number of objects detected in Google Street View Imagery and population estimates. We computed the coefficient of determination for (A) people (B) cars (C) trucks (D) trains.

• We **tested** the performance of our predictor in leave-one-out cross-validation using U.S. Census bureau estimations over years 2010 to 2016 (weighted by the corresponding proportions of imagery) as ground truth.

Mean absolute error: **21367 ± 28568**
Minimum absolute error: **20**
Maximum absolute error: **188988**
Root mean squared error: **35675**

Conclusions and Future Work

• The accuracy of a simple model for population estimation from GSV imagery is disappointingly poor, but could be improved by including additional features and assembling a larger dataset.

• Train, using transfer learning, our object detector to count other useful objects (windows, doors, recycling bins, etc.).

• Investigate the generalizability of a model trained on U.S. imagery to other countries.

References

[1] Robinson, C., Hohman, F., & Dilkina, B. (2017). A Deep Learning Approach for Population Estimation from-Satellite Imagery (Vol. 1996). Retrieved from http://arxiv.org/abs/1708.09086

[2] Gebru, T., Krause, J., Wang, Y., Chen, D., Deng, J., Aiden, E. L., & Fei-Fei, L. (2017). Using deep learning and Google Street View to estimate the demographic makeup of neighborhoods across the United States. Proceedings of the National Academy of Sciences of the United States of America, 114(50), 13108–13113.

[3] Rundle, A. G., Bader, M. D. M., Richards, C. A., Neckerman, K. M., & Teitler, J. O. (2011). Using Google Street View to Audit Neighborhood Environments. American Journal of Preventive Medicine, 40(1), 94–100.

[4] Dick, K., Charih, F., Dosso, Y. S., Russel, L., & Green, J. R. (2018). Systematic Street View Sampling. In Submitted to the 15th Conference on Computer and Robot Vision 2018 (pp. 1–8). Ottawa, ON.

[5] Lin, T., Ai, F., & Doll, P. (2008). Focal Loss for Dense Object Detection.