Estimating Building Age from Google Street View Images Using Deep Learning

Yan Li
Department of Infrastructure Engineering, University of Melbourne/Melbourne, Australia
li17@student.unimelb.edu.au

Yiqun Chen[1]
Department of Infrastructure Engineering, University of Melbourne/Melbourne, Australia
yiqun.c@unimelb.edu.au

Abbas Rajabifard
Department of Infrastructure Engineering, University of Melbourne/Melbourne, Australia
abbas.r@unimelb.edu.au

Kourosh Khoshelham
Department of Infrastructure Engineering, University of Melbourne/Melbourne, Australia
k.khoshelham@unimelb.edu.au

Mitko Aleksandrov
Department of Infrastructure Engineering, University of Melbourne/Melbourne, Australia
mitko.aleksandrov@unimelb.edu.au

Abstract

Building databases are a fundamental component of urban analysis. However such databases usually lack detailed attributes such as building age. With a large volume of building images being accessible online via API (such as Google Street View), as well as the fast development of image processing techniques such as deep learning, it becomes feasible to extract information from images to enrich building databases. This paper proposes a novel method to estimate building age based on the convolutional neural network for image features extraction and support vector machine for construction year regression. The contributions of this paper are two-fold: First, to our knowledge, this is the first attempt for estimating building age from images by using deep learning techniques. It provides new insight for planners to apply image processing and deep learning techniques for building database enrichment. Second, an image-base building age estimation framework is proposed which doesn't require information on building height, floor area, construction materials and therefore makes the analysis process simpler and more efficient.

2012 ACM Subject Classification Computing methodologies → Supervised learning by regression

Keywords and phrases Building database, deep learning, CNN, SVM, Google Street View

Digital Object Identifier 10.4230/LIPIcs.GIScience.2018.40

Category Short Paper

Funding The first author would like to acknowledge the financial support of China Scholarship Council (CSC).

---

[1] Corresponding author

10th International Conference on Geographic Information Science (GIScience 2018).
Editors: Stephan Winter, Amy Griffin, and Monika Sester; Article No. 40; pp. 40:1–40:7
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

1 Introduction

Building databases have been widely used for urban planning. New construction and renovation works require comprehensive building databases for analysis and decision-making. However, the application of such databases has been hampered by data integrity and accuracy issues [14].

Many image databases are available online. For example, Google Street View, which updates an online street image database periodically, provides advanced APIs for accessing building images by the location. Google street view based applications have been implemented in urban planning, such as estimating the demographic makeup of the cities [6], studying the relationships between city appearance and the health of its residents [5]. Another example is Google search, when appropriate keywords are provided, hundreds of relevant images will show in the results. Several public image classification databases are created from this data source, such as ImageNet [3] and CIFAR-10 [9].

Image processing using deep learning methods has shown great performance in many applications, such as image classification [10] and object segmentation [2]. Several popular deep learning methods have been developed for image analysis. Convolutional neural network (CNN) is the state-of-art method for feature extraction [10]. Support vector machine (SVM) has shown remarkable performance in regression problems, especially for high dimensional data [4]. Despite the rapid development of image processing techniques, building age estimation from images has not been studied in the research community. Existing methods such as [1] require additional building attributes (e.g. building height, floor area, etc.) for decision-making. Collecting these attributes is time-consuming and usually ends up with incomplete information. This paper proposes a novel method based on deep learning approach for direct building age estimation, using the CNN for image feature extraction and SVM for building age estimation.

The contributions of this research are two-fold:

- this is the first attempt for estimating building age from Google Street View images by using deep learning techniques. It provides new insight for planners to apply image processing and deep learning techniques for building database construction.
- the proposed image-based building age estimation framework is independent of building information, such as height, floor area, construction materials; therefore, it makes the analysis process simpler and more efficient.

2 Related work

2.1 Building age estimation

While building age is an important parameter in building specifications, the data is not always available or complete. Little research has been done for the building age estimation. [1] proposed an estimation method which adopts random forest regression and infers the building construction age from other attributes, such as ceiling height, footprint area, shape complexity and so on. However the accuracy of this method largely depends on the completeness of these attributes. This limitation motivates us to seek alternative solutions for overcoming the native incompleteness of existing database attributes. So looking for the available dataset is critical for our research and open data is ideal for this purpose. Several large Internet companies such as Google, Facebook, Instagram provide free APIs to access their image sources, in particular, Google Street View API provides house images based on a given location and hence it meets our requirement.

2.2 Image regression

This research treats the building age estimation as a regression problem. Image regression, which builds a regression model based on extracted image features, has shown the state-of-art performance in many tasks.

The general procedure of image regression contains two steps. The first step is to retrieve features from images, and the second step is to construct a regression model using the extracted features as inputs. The CNN [15, 11] is adopted in this research. Since the training of a CNN requires large training datasets and computing resources, many pre-trained models have been made publicly available. For example, the place365 dataset [16], trained by 8 million images, is used for scene recognition. Pre-trained CNN models are provided including AlexNet [10], ResNet18, ResNet50 [7], DenseNet161 [8], which are high performance CNN structures. As for the regression model, the support vector regression (SVR) can capture main features that characterize the algorithm (maximal margin). It is particularly suitable for high dimensional regression problems with a limited number of training samples.

3 Methodology

Our approach includes three major steps: data collection, feature extraction, and building age regression. First, the house images of each address are obtained from Google Street View API. Second, image features are extracted using a pre-trained CNN. At last, a SVR model is built by taking image feature vectors as inputs and building age as outputs.

3.1 Google Street View images download

Using the Google Street View Image API, we directly submit a list of addresses, for example, "172 Bouverie St, Carlton VIC 3053", and then store the retrieved house images locally. This process avoids the potential accuracy problems introduced by geocoding procedure and successfully obtains all the house images except invalid street addresses. As the images are shot from streets, they usually contain the target house in the mid as well as parts of adjacent buildings on two ends. We tune the API parameters to obtain the exact region of the target house. The parameters include heading, pitch (the horizontal and vertical rotation of the camera respectively) and fov (field of view, controlling the width of the street view images). Based on our experiments, setting heading as 180, pitch as 0, and fov as 50 degrees yields the best image results. In particular, the fov should be assigned appropriately because a wide view will introduce neighbour buildings and a narrow view will only capture partials of the target building. Each retrieved building image is in 600x400 pixels, which is the largest size that Google Street View API provides.

3.2 Feature extraction by Convolutional Neural Network

In this paper, we choose the largest scene recognition database and three pre-trained CNN models including AlexNet, ResNet and DenseNet for image feature extraction, as shown in Figure 1. These models are different in network structures. AlexNet won the 2012 Imagenet competition. Compared with modern network structures, AlexNet is simple and consists of 5 convolutional layers, maxpooling layers, drop-out layers and three fully-connected layers. It is specially designed for classification with 1000 categories. ResNet won the 2015 ImageNet and COCO competitions, and it allows for effectively training deeper neural networks. DenseNet, proposed in 2016, is based on the hypothesis that convolutional networks can be substantially

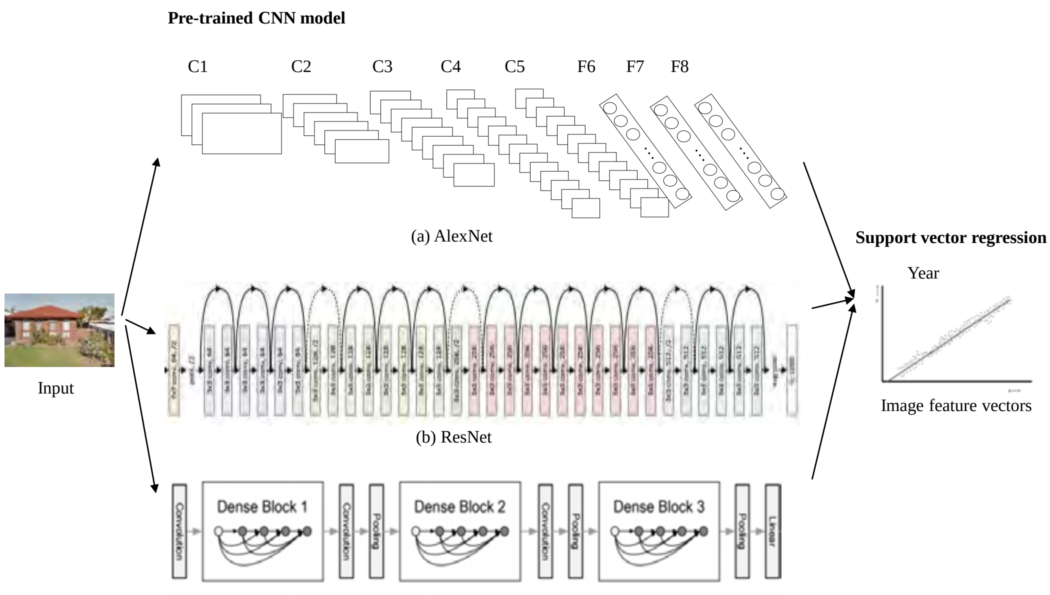

Figure 1 Different Convolutional neural networks are applied to extract image features. The convolutional networks are getting deeper from AlexNet to DenseNet.

deeper, more accurate and efficient to train if they build shorter connections between each layer and every other layer. All the experiments of extracting image features are implemented using the deep learning framework Pytorch [12].

3.3 Support Vector Regression

The support vector regression (SVR) [4] is advanced in high dimensionality space because SVR optimization doesn't depend on the dimensionality of the input space, and provides different kernel functions for the decision function. This research chose Scikit-learn library [13] to build the SVR model by taking image vectors as inputs and building age (construction year) as outputs. 80% of data are used to train the regression model and the best fit SVR model is decided according to the training data. Then we perform regression on test data based on the trained model.

4 Experiment results

4.1 Dataset

As shown in Figure 2, the North and West Metropolitan Region (NWMR) is chosen as the case study area, which is the most populous and diverse region extending from the Melbourne CBD to the outer northern and western suburbs in Victoria, Australia. It has 2981 square kilometres, 14 local government areas and around one third (33.1%) of the population of Victoria (2011 Census). The building attributes for NWMR are extracted from Valuer-General Victoria valuation dataset which contains the location, street address, zoning type, construction year, building material and valuation prices for both land and property across the entire Victoria. We assume that the building images from Google Street View for different zoning types may vary significantly and weigh down the model performance, hence

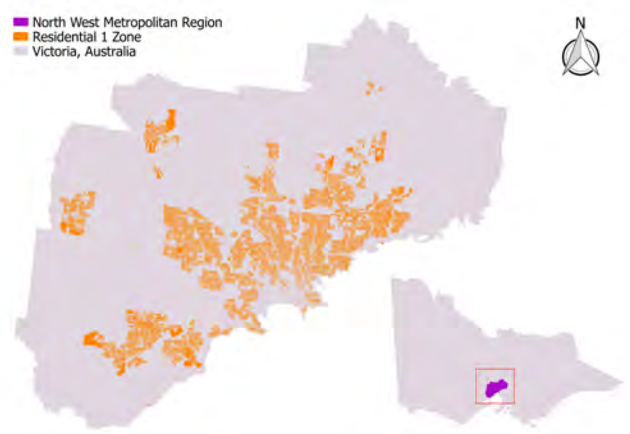

Figure 2 Case study area: North and West Metropolitan Region (NWMR), Victoria, Australia

Table 1 Accuracy of each CNN structure

CNN structure	AlexNet	ResNet18	ResNet50	DenseNet161
MAE	10.749	10.996	10.722	10.689
RMSE	12.210	12.423	12.154	12.121

the dataset is further narrowed down to Residential 1 Zone (R1Z) which contains 520,694 (69.5%) buildings in NWMR. It also should be noted that among these R1Z buildings, 21,830 (4.19%) of them miss construction year (i.e., building age) information. The key motivation of this work is to estimate the missing values.

4.2 Accuracy

For regression models, two evaluation metrics are widely used for performance evaluation: mean absolute error (MAE) and root mean squared error (RMSE), both indicate the error of prediction results. MAE is the average over the test sample of the absolute differences between the prediction and the actual observation where all individual differences have equal weights. RMSE is the square root of the average of squared differences between the prediction and the actual observation. Since the errors are squared before they are averaged, RMSE is more useful particularly when large errors are undesirable.

Table 1 summarises the estimation performance of different CNN structures. DenseNet161 shows the best performance among them, and it confirms that deeper CNN structure tends to yield more accurate results[8]. Figure 3 shows the distribution of errors. Around 15% samples have less than 5 years error. Most samples, about 25%, have about 10 years error.

4.3 Findings

The changing of building fashions allows inspectors to roughly determine when buildings are constructed, based on their appearances, materials, components and styles. Inspired by this idea, we explore the feasibility of teaching a machine to estimate the building age by reading housing images and learning the styles. We list house samples in the same age range in Figure 4 and find some interesting patterns. Clearly, it can be observed that more recently constructed buildings tend to have newer facades, and houses are getting more complex both

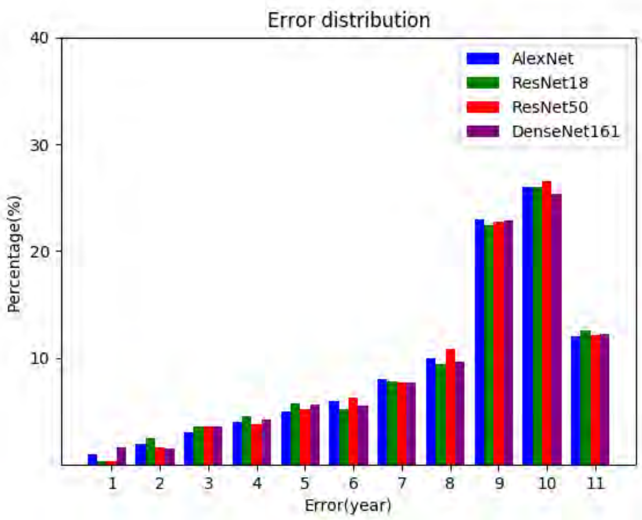

Figure 3 Error distribution of each CNN structure

in horizontal and vertical space. Duplex houses also prevail in recent decades. Exterior wall materials have also changed over time. Before 2000, newly built houses had wood or brick exteriors; while after that, new houses start to use vinyl siding. These features are hidden in images and could be learned and extracted by the CNN models and then passed to our SVR model for the regression analysis.

5 Conclusions and future work

In this study, a novel approach for direct estimation of building age from Google Street View images is proposed, implemented and tested. The algorithm consists of three major steps: Google Street View images download, image features extraction and building age estimation. Results for the North and West Metropolitan Region of Victoria show that building age estimation can be accurately predicted with deeper convolutional neural networks.

References

1 Filip Biljecki and Maximilian Sindram. Estimating building age with 3d gis. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, IV-4/W5, 2017.

2 Ali Borji, Ming-Ming Cheng, Huaizu Jiang, and Jia Li. Salient object detection: A benchmark. *IEEE Transactions on Image Processing*, 24(12):5706–5722, 2015.

3 J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. ImageNet: A Large-Scale Hierarchical Image Database. In *CVPR09*, 2009.

4 Harris Drucker, Christopher JC Burges, Linda Kaufman, Alex J Smola, and Vladimir Vapnik. Support vector regression machines. In *Advances in neural information processing systems*, pages 155–161, 1997.

5 Abhimanyu Dubey, Nikhil Naik, Devi Parikh, Ramesh Raskar, and César A Hidalgo. Deep learning the city: Quantifying urban perception at a global scale. In *European Conference on Computer Vision*, pages 196–212. Springer, 2016.

6 Timnit Gebru, Jonathan Krause, Yilun Wang, Duyun Chen, Jia Deng, Erez Lieberman Aiden, and Li Fei-Fei. Using deep learning and google street view to estimate the demo-

Figure 4 Samples of buildings with estimated construction year

graphic makeup of neighborhoods across the united states. *Proceedings of the National Academy of Sciences*, page 201700035, 2017.

7 Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

8 Gao Huang, Zhuang Liu, Laurens van der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017.

9 Alex Krizhevsky and Geoffrey Hinton. Learning multiple layers of features from tiny images. Technical report, University of Toronto, 2009.

10 Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.

11 Yan Li, Majid Sarvi, Kourosh Khoshelham, and Milad Haghani. Real-time level-of-service maps generation from cctv videos. In *Transportation Research Board 97th Annual Meeting*, 2018. URL: `https://trid.trb.org/view/1497380`.

12 Adam Paszke, Sam Gross, Soumith Chintala, Gregory Chanan, Edward Yang, Zachary DeVito, Zeming Lin, Alban Desmaison, Luca Antiga, and Adam Lerer. Automatic differentiation in pytorch, 2017. URL: `https://openreview.net/forum?id=BJJsrmfCZ`.

13 F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

14 Abbas Rajabifard. *Spatial data infrastructure*. International Federation of Surveyors (FIG), 2012.

15 Jürgen Schmidhuber. Deep learning in neural networks: An overview. *Neural networks*, 61:85–117, 2015.

16 Bolei Zhou, Agata Lapedriza, Aditya Khosla, Aude Oliva, and Antonio Torralba. Places: A 10 million image database for scene recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017.