

Predicting Urbanization from Daytime Satellite Imagery based on Descriptive Statistics

Donghyun Ahn^{1,2} Sungwon Han^{1,2} Hyunji Cha^{1,3} Meeyoung Cha^{1,2}

¹Data Science Group, Institute for Basic Science, Daejeon, South Korea

²School of Computing, KAIST, Daejeon, South Korea

³Software Graduate Program, KAIST, Daejeon, South Korea

ABSTRACT

Cities are overgrowing worldwide, and this urbanization process is sometimes faster than the development of the necessary infrastructure needed for the urban population. Unfortunately, tracking urban growth has been difficult due to the lack of data, particularly for rapidly evolving developing countries. This paper presents a statistical model that quantifies the urbanized features from daytime satellite imagery. The model can process high dimensional vector spaces efficiently and expresses the degree of urbanization in scalar value. The model predicts urban-related demographics, such as population count and spending per capita, with high R-squared values. Such a model can help solve multiple urbanization problems in the context of sustainable development goals (SDG).

1. Introduction

By 2050, more than two-thirds of the world's population will live in urban areas, up from about 54% today to 68% [6]. Urbanization is faster in developing countries, which expects a higher population density as a massive population in rural areas and suburbs migrate to urban cities. By 2050, this translates to the occupancy of 416 million urban dwellers in India and 189 million in Nigeria alone. Understanding the patterns of urban growth and measuring the progression of regional urbanization remains a global challenge, albeit this task is critical for predicting future population changes and for identifying the potential for the development of the region. In particular, analyzing urban growth provides a data science method for setting up and diagnosing national-level policies, regional problem assessments, and market research.

Unlike other socio-economic measures that largely depend on infrastructures for data collection, fine-grained satellite images are readily available at a planetary scale and over time and hence can be considered an economic resource for analysis. For example, the ArcGIS Image Tile REST API¹ offers cloud-free high-resolution satellite image tile data. A satellite image is in units of zoom levels, z (0~18), which defines the coverage of the region. Each tile has a resolution of $156,412\text{m}/2^z$ pixel, so the collectible satellite images have a maximum resolution of 0.596m/pixel. Satellite image data sources like the one provided by ArcGIS exist for many parts of the world, including rapidly evolving developing regions. The challenge, then, lies at developing artificial intelligence-based techniques to gather, clean, and analyze the massive amounts of satellite images to gain data insights on the level of urban land use. Data insights on urbanization will be critical to tracking economic activities (e.g., purchasing power) and sustainable development (e.g., energy systems).

This paper presents a novel model that predicts the spatiotemporal patterns of urbanization

¹ArcGIS Image Tile REST API Documentation, <https://developers.arcgis.com/rest/services-reference/image-tile.htm>

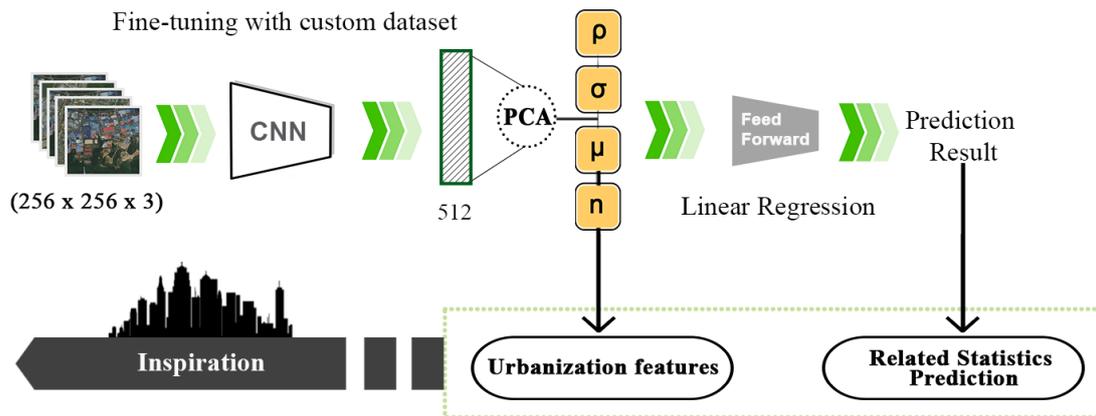


Figure 1. High-level architecture of the proposed model, extracting features of urbanization patterns from satellite imagery

based on embedded space statistics from satellite images gathered from target regions. The constructed model comprises a convolutional neural network (CNN) that is known to perform well in image processing [1, 5]. The crux of the model is at its ability to jointly utilize a large number of satellite images, by capturing their mean and standard deviation statistics, and projecting the high-dimensional data to a lower-dimension picture, typically using 2 or 3 projected dimensions. This simple yet elegant process — depicted in Figure 1 — could well distinguish the traits of urban and rural land usages, as well as identify local urbanization traits using embedded features extracted from satellite imagery.

For validation, the model training utilized publicly available satellite images collected on specific countries. Here, this research reports results on the South Korea data and its urbanized demographics. The choice of South Korea is of particular interest because it is one of the planet’s most densely populated countries (ranking 28th). The country’s capital, Seoul, exhibits high skylines and advanced urban land use, holding nearly one-fifth of its population, whereas many rural parts are known to have less benefit from the urban growth and occupied by the elderly. Hence, the various set of demographics to predict—population composition, education, and spending power— have wide variance within the country. Despite the skewed data, the model could predict certain demographic features with high accuracy. For instance, regression to population density over South Korean districts showed results with an R-squared value of 0.9418.

Although preliminary, this research demonstrates that the readily available vibrant satellite images and the latest deep learning methods presented in this paper could be extended to capture urbanization processes in various countries and also quantify a wide range of demographics. The proposed model has strengths, mainly in the context of sustainable development goals (SDG) for developing countries. Demographic data had been collected via either offline or online surveys, which are costly and time-consuming. Unlike the existing survey method, satellite images periodically taken in real-time are made available on the Internet. By leveraging data insights, countries and regional governments can better decide on policies that promote sustainable industrialization and foster innovation.



Figure 2. Our space embedding could efficiently identify rural (blue dots) and urban (red dots) areas. Randomly chosen images from the urban classification, shown on the right hand side, could further represent the varying degree of urbanization. Population density increases from images (1) to (4), which were chosen along the embedded space on the plot.

2. Method for training embedding via transfer learning

Estimating the degree of urbanization is hard mainly due to ambiguity in its meaning. Even if there exist guidelines on urbanization degrees, such as population density or purchasing power, it is hard to set novel standards for quantifying them systematically. For example, there would be many regions that cannot be associated with a single label. Therefore, a statistical model, including machine learning with no target value, has difficulty in analyzing the urbanization process. Therefore, this research proposes the adoption of a new technique, namely, *transfer learning*.

Transfer learning is a machine learning technique that concentrates on the knowledge transfer from different yet related original problems [8]. This technique applies to cases where the data of the main problem is not large enough to train deep neural networks, which often require million-scale training instances. By utilizing a large dataset that is readily available, the so-called proxy task, knowledge from the data is passed to a neural network to help solve a target problem with good initial weight points [11]. Moreover, transfer learning can be applied to construct feature extractors with knowledge from proxy data [4, 10]. Depending on what knowledge one embeds in the data, the features extracted from neural networks would vary. Urbanization itself is an ambiguous concept, yet one may define it with various aspects. This research utilizes transfer learning to train a neural network and to extract the features of urbanization from satellite images.

Tuning convolutional neural networks (CNNs) with a set of finely configured proxy data with relevant knowledge produces embedded vector representations of urbanization. The satellite image was pretrained into 512 dimensions with ResNet-18, a deep neural network that is known to show high performance on image classification tasks [3]. The model projects this high-dimensional data to a lower-dimension picture—typically utilizing 2 or 3 projected dimensions. This projection uses principal component analysis (PCA) and the optimal number of dimensions varied by training instances.

Experiments were conducted to verify whether the proposed technique is valid. Four human annotators made a custom dataset with 1000 randomly selected images from South Korea for proxy. Four independent human annotators were recruited to label whether a satellite image represents rural, urban, or uninhabited areas (e.g., forests). Then, we trained the CNN-based classifiers with the custom dataset. Urban and rural images are reduced to embedded vectors by a feature extractor. The extracted features, projected over 3-dimensional space through PCA, are presented in Figure 2. This result demonstrates that images from urban (red dots) and rural (blue dots) are well-separated on embedded space. Urban images were divided into four subgroups according to the vector size value in the extending direction. As one can

Table 1. Linear regression result on urbanize-related demographics

Variable ID	Description	R-squared
<i>total</i>	Total population count	0.7988
<i>density</i>	Population per square kilometer	0.9418
<i>edu.phd</i>	Population count by Ph.D. degree	0.7492
<i>purchase.total</i>	Total purchasing power per household	0.8087
<i>purchase.capita</i>	Purchasing power per capita	0.6506
<i>sp.alcohol</i>	Spending on alcoholic beverage per capita	0.6319

see on the right side of the figure, there exist significant differences among subgroups correlated with urbanization. If the size of vectors in urban data direction becomes larger, images are more likely to capture areas with high density. This result demonstrates that the fine-tuned network well reflects the characteristics of rural and urban in this experiment.

3. Results on features with descriptive statistics

The number of satellite images differs from one municipal region to another. Hence, there exist a different number of extracted features of urbanization per municipal districts. Then, to obtain a fixed length feature across all areas without information loss, one needs a novel way to treat data. The crux of the proposed model is its ability to jointly utilize a large number of satellite images by capturing their mean and standard deviation statistics.

Let us assume that satellite imagery from various municipal regions follows a particular distribution. Those distributions will vary according to characteristics of the corresponding areas (e.g., degree of urbanization). By inferring parameters from distributions that capture traits of cities, one can obtain useful information that reflects a tendency across urbanization. Critical parameters of the observed distributions were extracted by summarizing observations through descriptive statistics, which is one of the major branches in sample analysis [9]. Descriptive statistics concentrate on summarizing given data by representing central tendency (mean, median, mode), dispersion (variance, standard deviation, skews), and association (chi-square, correlation). Measures of these properties can explain the summarized characteristics of the entire regions covered by satellite images to a great extent. Among these, the method adopts the mean and standard deviation statistics of satellite images representing the same municipal region.

We further tried to verify the usefulness of the features once more through urbanize-related demographics predictions. Verification involved 69 demographics collected by the 2018 Michael Bauer Research² on South Korean Districts. It reports various demographics such as population itself, population by age/education level, income, and spending of a target area. Linear regression was conducted on log-scaled demographics over all the districts. Table 1 displays six remarkable results among the 69 demographics attributes. For example, the population per square kilometer shows the highest R-squared value. Spending on alcoholic beverage per capita could be predicted, yet with a lower R-squared value.

²Michael Bauer Research GmbH, <http://www.english.mb-research.de/>

4. Concluding remarks

Several recent studies have proposed utilizing satellite images to understand urbanization [2,7]. This research demonstrated how urbanization, as well as advanced statistics, can be captured via deep neural networks. Unlike previous approaches that only use sample images or nighttime satellite images, the proposed method can efficiently handle massive amounts of high-resolution data. This method utilizes *all* satellite images on municipal regions by adopting descriptive statistics. This approach is lightweight and does not require far more computing power than existing methods. By using all images available, the process can reduce any noise from random selection and allow one to recognize high-level variation on the entire target region.

The presented methodology can be extended and adopted by national and regional governments to estimate the socio-economic growth of urban and rural areas. This method could benefit developing countries, particularly, those that lack the infrastructure that can be used to monitor the rapid urbanization process. Results indicate that the extracted features from fine-tuned neural networks could represent the urbanization traits well (see Figure 2), and their sample descriptive statistics were sufficient for the summary. The embedding was also confirmed to be high-quality, based on urbanize-related demographics predictions. Regression models were built on features, reported from 0.6319 to 0.9418 R-squared values, for predicting socio-economic measures of the studied areas.

References

- [1] CIRESAN, D. C., MEIER, U., MASCI, J., GAMBARDILLA, L. M., AND SCHMIDHUBER, J. Flexible, high performance convolutional neural networks for image classification. In *Twenty-Second International Joint Conference on Artificial Intelligence* (2011).
- [2] DOUPE, P., BRUZELIUS, E., FAGHMOUS, J., AND RUCHMAN, S. G. Equitable development through deep learning: The case of sub-national population density estimation. In *proc. of the ACM Annual Symposium on Computing for Development* (2016), p. 6.
- [3] HE, K., ZHANG, X., REN, S., AND SUN, J. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (2016), pp. 770–778.
- [4] JEAN, N., BURKE, M., XIE, M., DAVIS, W. M., LOBELL, D. B., AND ERMON, S. Combining satellite imagery and machine learning to predict poverty. *Science* 353, 6301 (2016), 790–794.
- [5] LONG, J., SHELHAMER, E., AND DARRELL, T. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (2015), pp. 3431–3440.
- [6] NATIONS, U. 2018 revision of world urbanization prospects, 2018.
- [7] NÚÑEZ, J. M., MEDINA, S., ÁVILA, G., AND MONTEJANO, J. High-resolution satellite imagery classification for urban form detection. In *Satellite Information Classification and Interpretation*. IntechOpen, 2019.
- [8] PAN, S. J., AND YANG, Q. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering* 22, 10 (2010), 1345–1359.
- [9] PECK, R., OLSEN, C., AND DEVORE, J. L. *Introduction to statistics and data analysis*. Cengage Learning, 2015.
- [10] XIE, M., JEAN, N., BURKE, M., LOBELL, D., AND ERMON, S. Transfer learning from deep features for remote sensing and poverty mapping. In *Thirtieth AAAI Conference on Artificial Intelligence* (2016).
- [11] YOSINSKI, J., CLUNE, J., BENGIO, Y., AND LIPSON, H. How transferable are features in deep neural networks? In *Advances in neural information processing systems* (2014), pp. 3320–3328.