WORKING PAPER: PLEASE DO NOT CITE

Using Convolutional Neural Networks and Very-High Resolution Satellite Imagery to Classify New Delhi into Planned vs. Informal Settlements

Bob Bell, bob.bell@un.org, UNCTAD

Rajesh Veeraraghavan, rajesh.veera@georgetown.edu; Georgetown University

ABSTRACT

The main idea of the paper is that convolutional neural networks can be applied to very high-resolution satellite imagery in order to classify New Delhi into formal (planned colony) vs. informal settlements (Jhuggi Jhopri Clusters). We show that very high-resolution satellite imagery along with convolutional neural networks can achieve a classification accuracy of 95.81%. We find that pre-trained deep learning models for computer vision trained on standard image datasets can be effective for classification of informal settlements using satellite imagery, even when there is not a significant amount of training data. Deep learning models can learn image features without hand-crafted features and when coupled with the proliferation of cloud-based computer vision services could democratize the analysis of satellite imagery for humanitarian and developmental purposes.

•KEYWORDS

machine learning; slum classifier; informal settlements; detecting slum boundaries; convolutional neural network

1 Introduction

By 2030, Indian cities are projected to be home to over 630 million residents but there are clear signs that governance capacity and patterns of public investment are inadequate [1]. While cities have gotten a lot of recent attention, the discussion has been dominated by understanding questions of planned settlements, there has been relatively very little that we know about the informal settlements in Delhi. A primary reason for the increased attention in planned settlements is that there is more information available on the former and we don't even have an adequate picture of the informal settlements [2]. India's government has proposed a nationwide program to build 100 smart cities and the US has committed significant resources to supporting this initiative [3]. The Indian government defines a smart city as a city "equipped with basic infrastructure to provide a decent quality of life, and a clean and sustainable environment through the application of some smart solutions" [3]. It is estimated that over the next few years India's infrastructure investment will be in the $1.5 to $2 trillion range. Yet, by its own admission, the Indian government and city governments have a very limited understanding of the spatial distribution of basic infrastructure and its socio-economic effects. For example, despite being the best resourced city in India, three quarters of Delhi's population live in "unplanned settlements" that have limited or no access to basic services such as piped water and sanitation, few if any public facilities and are poorly connected to the city's transport grid [2].

There is a nascent quantitative literature on urban India that has drawn attention to the spatial dynamics of exclusion [4][5][6]. These studies have pointed to a clear statistical relationship between spatial location and service quality. On the other hand, the Cities of Delhi (COD) project uses qualitative methods to carefully document the differentiation of social citizenship across Delhi's unplanned settlements [7]. In these unplanned settlements which represent roughly half of the city's population, basic service delivery (water, sanitation, garbage removal and access to transport) is extremely poor and lags far behind levels prevailing in planned settlements. The main finding of COD is that independently of class and other social characteristics, where a citizen lives in Delhi determines the level of basic services to which he or she is entitled. This has significant negative spillovers, since these services are essential to supporting core capabilities such as health, education, and economic opportunity. Differentiated citizenship, in other words, sustains and amplifies inequality by excluding residents from capability-enhancing public services. What is more, the extent of this exclusion seems to be growing.

Despite continuous "reforms", radically increased levels of infrastructural investment and a competitive political environment, much of the city remains unplanned and underserviced. Data on settlements types is unreliable and vastly underestimates the informal settlement populations (COD). The COD project does provide a more granular picture of service delivery across informal settlements and identifies some mechanisms, but is limited in its generalizability based as it is on case studies of 16 out of an estimated 2,000 unplanned settlements. As the COD project documents, state agencies are complicit in maintaining the illegality or even invisibility of unplanned settlements [7].

Further, because of the underdeveloped statistical capacity in many countries in the global south, especially for these urban informal settlements, it is important to provide alternative and complementary indicators for helping local governments, civil society, identify segments of urban areas and track public services accordingly. There is a promising path set out by work done by Blumenstock on using mobile phone data to make statistical inferences on estimates of poverty [8].

We believe that satellite imagery and machine learning methods can help developing countries, in particular, estimate the areas of cities with slums and informal settlements. Furthermore, these spatial predictions can be used to infer a number of characteristics, including the proportion of the urban population living in slums and informal settlements and access to basic services and infrastructure.

Remote sensing imagery and machine learning have the potential to radically reduce the cost of getting population estimates in slums so that public authorities can make evidence-informed decisions for slum upgrading projects. Our approach, allows us to make it easy to bootstrap data collection where such systematic efforts by the state does not exist, as well as to go beyond state's own data, and to create the possibility of social audits by citizen groups.

Comprehensive surveys would be prohibitively expensive, and we would not have any possibility to dynamically update data. On the other hand, traditional image processing and conventional computer vision methods could potentially be used to classify some settlements where spatial features can be easily seen from satellite imagery but rely on extensive domain expertise and hand-crafted feature engineering.

The benefit of using convolutional neural networks (CNNs) is that the specification of features - whether specific image filters or image processing techniques - is not required. Rather, CNNs can automatically discover possible spatial features, if they exist, given enough training data and with appropriate architectures.

This paper employs a binary classification task to differentiate informal settlements - the *Jhuggi Jhopri Cluster (JJC)*, which represents a legal category that represents the marginalized group - versus planned colonies. We take advantage of the data now provided by commercial satellite providers as well as the use of new machine learning approaches applied to computer vision [24]

The use of machine learning to detect informal settlements is an emerging area of research [12]. As Gechter point out in their analysis this is an extremely high-dimensional problem [17]. Existing approaches to solving this high-dimensionality problem involve pre-specified procedures that look for features like vegetation index, building density, texture etc. [12]. Despite this, the accuracy of these studies hovered around 70-81% [9].

Recent approaches are increasingly incorporating machine learning for the automated identification of features. For example, CNN's have been used to classify informal settlements in Delhi and Mumbai using fully convolutional networks (FCN's). Specifically for Delhi, the overall accuracy is reported as: 88.41% (pre-trained FCNs), 81.12% (transfer learned FCNs), and 93.07% (fine-tuned FCNs that leverage models trained on one city to learn last layers for another city). [16]

World Bank conducted a study on Poverty from Space to predict poverty rates [13]. Their study combined CNN's for car and roof detection but then incorporated additional feature engineering (PanTex, HOG, SIFT, etc) into a larger classifier.

Datasets and Data Preprocessing

We procured very high resolution satellite imagery covering 893 sq. km. of 4-band (R/G/B/NIR1) pan-sharpened and orthorectified imagery with 31 cm resolution from DigitalGlobe. Pre-processing of imagery included pan-sharpening and ortho-rectification. We acquired imagery from 2015 and 2016, corresponding to scenes with minimal (<5 %) cloud cover.

Shapefiles on JJCs and planned colonies (KML files) was collected by the current Delhi Government political party [*Aam Aadmi Party*] over the 2015-16 period. For this initial analysis, only RGB bands were included from the original 4-band multispectral data. The vector shapefiles were reprojected to the coordinate reference system of the GeoTIFF raster files (EPSG: 32643). The individual polygon features of the shapefile were masked onto the GeoTIFF files, generating separate GeoTIFF images for each feature but retaining the same metadata as the original raster files.
In order to create image chips of equal dimension for machine learning, the individual GeoTIFF files were tiled into blocks of 80 by 80 pixels that excluded image areas with no data. Given the non-uniform shape of planned colonies and JJCs, this process inevitably excluded some valid data from the original raster files, esp. JJCs which have irregular and narrow shapes. This resulted in 776 image chips for JJCs. To ensure a balanced

dataset, 776 image chips were randomly chosen from the 180,702 image chips available for the planned colony class.

Details of Machine Learning

These image chips were then fed into a PyTorch deep learning (0.3.0) pipeline based on the fast.ai Python library (0.7.0) with the following steps:

We defined 80% training set and 20% validation set, with error function as negative log likelihood loss, given the binary classification task. This results in a validation set of 310 image chips on which to assess accuracy. For the algorithm, we set the image size to 80x80x3 (R/G/B).

We selected ResNet architecture because of its excellent performance on image classification tasks for the 2015 ILSVRC and COCO competitions [18].[1] We trained the model on a number of different ResNet architectures (ResNet-18/34/50/101/152) as well as VGG-19 [19], which was utilized in a similar study [16].

To improve training performance, we utilized the CuDNN deep learning package which provides accelerated functions for working with Nvidia GPUs.

Using a pre-trained ResNet model originally trained on ImageNet (1.2 million images and 1000 classes), we trained for 3 epochs with a learning rate of 1e-2. This involved freezing all convolutional layers and only learning the last fully-connected (linear) layer for this specific classification task.

In addition to the pre-trained model, we extended the training to include the following steps: (1) identified an optimal learning rate, confirming that 1e-2 is optimal for model performance [20]; (2) data augmentation through top-down transforms as well as random zooming at a scale of 1.1; (3) apply stochastic gradient descent with restarts as the optimizer [21], a variant of learning rate annealing, to encourage our model to find parts of the weight space that are both accurate and stable; (4) differential learning rate annealing, where we use different learning rates for different layers, such that later layers have bigger learning rates than earlier layers, which typically have more general-purpose features [22]; (5) test time augmentation, which is data augmentation at inference time, which makes predictions not only on images in validation set but also on randomly augmented versions of them as well. This additional learning was also done with 3 epochs over the training set.

Total training time for all pre-trained models was approximately 2 hours and 15 minutes on a cloud GPU machine.[2] This cost approximately $2.70 on a GPU-enabled cloud computing platform.[3]

Results and Discussion

We found that we were able to classify informal settlements (JJCs) and planned colonies with a maximum accuracy of 91.61% using pre-trained models and 95.81% using the augmented pre-trained model (Table 1). Our work outperforms existing approaches using CNNs for slum detection.

The accuracy of a pre-trained model that only trained the last (fully-connected linear) layer typically increased with more convolutional layers, with the exception of ResNet-152. However, with the additional training steps (learning rate finder, data augmentation,

[1] See http://image-net.org/challenges/LSVRC/2015/ and http://mscoco.org/dataset/#detections-challenge2015.

[2] Tesla K80 with 12 GB memory, 61GB Ram, 100 GB SSD

[3] https://www.floydhub.com

stochastic gradient descent with restarts, differential learning annealing, and test-time augmentation), the accuracy of ResNet-18 matched or exceeded that of other models. This possibly implies that data augmentation techniques, in addition to the other techniques specified in the methods section, can improve classifier accuracy, regardless of the number of convolutional layers.

Table 1

Accuracy based on ML architecture

Architecture	Pre-trained model	Augmented Pre-trained model
ResNet-18	87.74%	**95.81%**
ResNet-34	89.03%	94.52%
ResNet-50	90.65%	**95.81%**
ResNet-101	**91.61%**	94.52%
ResNet-152	90.00%	94.84%
VGG-19	83.23%	95.48%

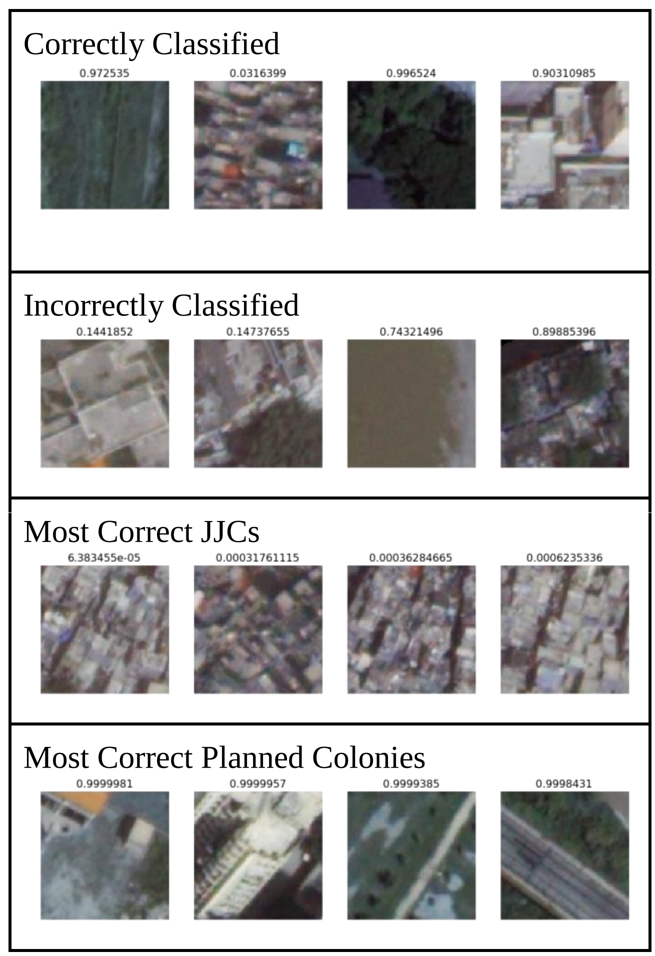
To understand the details of our classification, we present in Figure 1 both the correct and incorrect classification examples from the pre-trained classifier. In our approach, for the binary classifier, any probability above 0.5 had a predicted class of planned colony while a probability below 0.5 has predicted class of JJC.

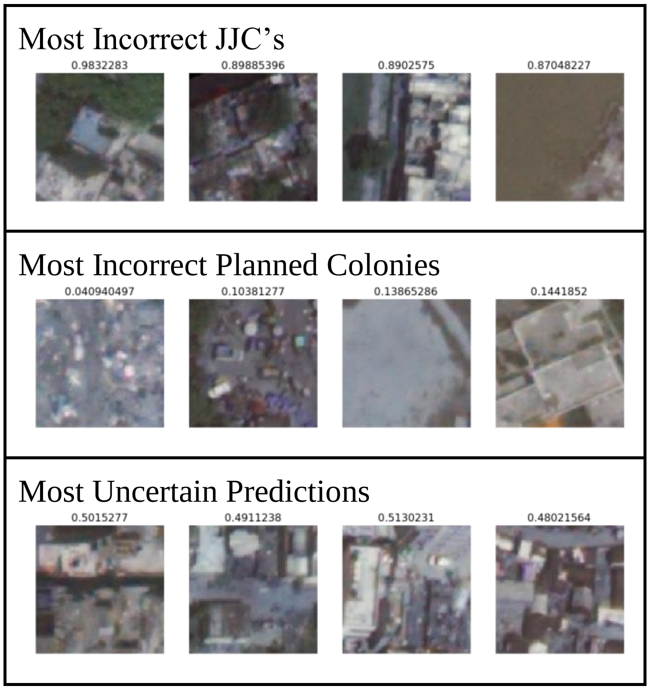
As Figure 1 shows, in the correctly classified examples, the images with irregular building structures are classified as JJCs while images with vegetation or rectangular building structures are classified as planned colony.

We noticed based on visual inspection that among examples of incorrect classification, a JJC was misclassified as a planned colony if vegetation was present in the image. On the other hand, an image of planned colony was misclassified as a JJC if it contained irregular shaped structures (e.g., presence of randomly parked cars or unclear building structures). Finally, the most uncertain predictions had a combination of rectangular and irregularly shaped building structures.

Figure 1

Illustration of the correct and incorrect classification examples from ResNet-34 pretrained model

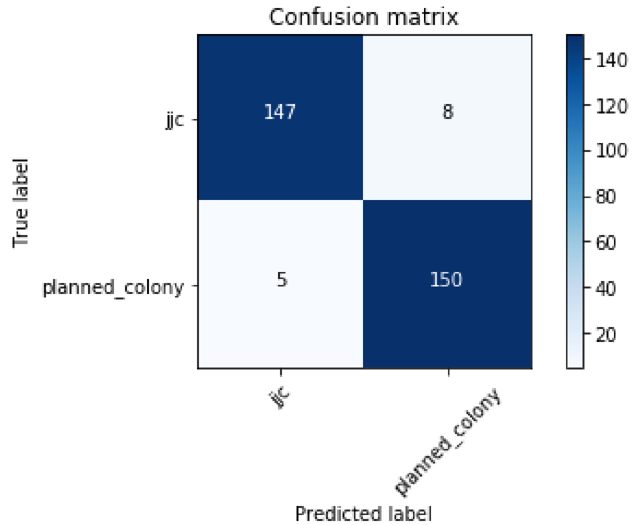




We had a test set of 310 images to predict our classification. We present the confusion matrix (figure 2), which represents how well the augmented ResNet model did in classifying planned and JJCs.

Figure 2

Confusion Matrix for Augmented ResNet-18



Conclusions and Future Work

We built a classifier to automatically detect slums from planned colonies. Our work outperforms existing approaches using CNNs for slum detection. Our classification was 95.81% using the augmented pre-trained model. We have shown that automated feature engineering techniques like CNN's, along with very high-resolution satellite imagery, can classify informal settlements which exhibit significant spatial, topographical, and socioeconomic heterogeneity.

We would like to extend this work beyond New Delhi and provide a general-purpose slum classifier for classifying informal settlements, particularly in cities in developing countries.

For this specific analysis, additional work can be conducted to potentially improve the accuracy of the classifier, including:

- Different neural network architectures with different numbers of layers (e.g., ResNext, Inception, GoogleNet, SqueezeNet, DenseNet)

- Additional forms of data augmentation (e.g., reflection padding and sliding window).
- Neural network frameworks (PyTorch 1.0) and associated libraries (fastai 1.0) which have additional neural network models and functionality.
- Explore different image tile sizes to identify optimal size that maximizes training set size and without sacrificing significant classifier accuracy.

We could also incorporate features typically characteristics of slum morphology, including: building size, roof material, absence of roads, irregular roads, lack of vegetation and open spaces, density, irregularly shaped settlements, texture, locality, and other factors [23]. Like the World Bank study, we could use CNNs for object detection and use those learned objects as features in a second stage of classification for slum classification.

The compute costs for training the classifier were minimal ($3), suggesting that as more low-cost and easy-to-use cloud services are provided for automated computer vision tasks, this can democratize the use of machine learning for international development and humanitarian relief.

We chose to use very high-resolution satellite imagery as a test case to validate the idea that slums can be detected using satellite imagery. However, the cost of imagery is still prohibitively expensive. In order to make this analysis truly accessible for pro-poor applications like slum detection would require that commercial satellite imagery providers and government agencies with satellite imagery to release such imagery at low or no cost. This is taking place to a limited extent for disaster relief.[4]

In the near-term, it would be ideal to extend this classification work with medium-resolution satellite imagery that is freely or cheaply available (e.g., from Planet, NASA Landsat, and ESA Sentinel). This holds the most promise for reproducibility and generalizability to other regions and reduces dependence on commercial satellite imagery providers.

REFERENCES

[1] http://www.un.org/en/events/citiesday/assets/pdf/the_worlds_cities_in_2018_data_booklet.pdf

[2] Bhan, G., & Jana, A. (2015). Reading Spatial Inequality in Urban India.Economic & Political Weekly,50(22), 49.

[3] http://smartcities.gov.in/content/

[4] Sidhwani, Pranav. 2015. "Spatial Inequality in Big Indian Cities." Economic and Political Weekly L(20): 55–62.

[5] Singh, Gayatri. 2014. Freedom to Move, Barriers to Stay: An Examination of Rural Migrants' Urban Transition in Contemporary India. PhD Dissertation, Department of Sociology, Brown University.

[6] Vithayathil, Trina, and Gayatri Singh. 2012. "Spaces of Discrimination." Economic and Political Weekly 47(37): 60–66.

[7] COD, 2015. Exclusion, Informality, and Predation in the Cities of Delhi, August 2015. http://www.cprindia.org/sites/default/files/working_papers/CoD_Overview.pdf

[8] Blumenstock, J. E. (2016). Fighting poverty with data. *Science*, *353*(6301), 753-754.

[9] Schmitt, A., Sieg, T., Wurm, M., & Taubenböck, H. (2018). Investigation on the

[4] https://www.digitalglobe.com/ecosystem/open-data

separability of slums by multi-aspect TerraSAR-X dual-co-polarized high resolution spotlight images based on the multi-scale evaluation of local distributions. *International journal of applied earth observation and geoinformation*, *64*, 181-198.

[10] Kleynhans, W., Salmon, B. P., & Olivier, J. C. (2015). Detecting settlement expansion in South Africa using a hyper-temporal SAR change detection approach. *International Journal of Applied Earth Observation and Geoinformation*, *42*, 142-149.

[11] Engstrom, R., Sandborn, A., Yu, Q., Burgdorfer, J., Stow, D., Weeks, J., & Graesser, J. (2015, March). Mapping slums using spatial features in Accra, Ghana. In *2015 Joint Urban Remote Sensing Event (JURSE)* (pp. 1-4). IEEE.

[12] Kuffer, M., Pfeffer, K., & Sliuzas, R. (2016). Slums from space—15 years of slum mapping using remote sensing. *Remote Sensing*, *8*(6), 455.

[13] Engstrom, Ryan; Hersh, Jonathan; Newhouse, David. 2017. Poverty from Space : Using High-Resolution Satellite Imagery for Estimating Economic Well-Being. Policy Research Working Paper;No. 8284. World Bank, Washington, DC.

[14] Heller, P., Mukhopadhyay, P., Banda, S., & Sheikh, S. (2015). Exclusion, informality, and predation in the cities of Delhi. *Centre for Policy Research*.

[15] Kuffer, M., Pfeffer, K., & Sliuzas, R. (2016). Slums from space—15 years of slum mapping using remote sensing. *Remote Sensing*, *8*(6), 455.

[16] Stark, T. (2018) Using Deep Convolutional Neural Networks for the Identification of Informal Settlements to Improve a Sustainable Development in Urban Environments. Master's Thesis Thesis Technische Universitut Munchen Photogrammetrie und Fernerkundung.

[17] Gechter M and Tsivanidis N. The Welfare Consequences of Formalizing Developing Country Cities: Evidence from the Mumbai Mills Redevelopment. Preliminary Draft available at: https://www.dropbox.com/s/srp8z3u5jlw9xi1/MumbaiMills.pdf.

[18] He, K., Zhang, X., Ren, S. and Sun, J., 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778). https://arxiv.org/abs/1512.03385

[19] Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556. https://arxiv.org/abs/1409.1556

[20] Smith, L. N. (2017, March). Cyclical learning rates for training neural networks. In *2017 IEEE Winter Conference on Applications of Computer Vision (WACV)* (pp. 464-472). IEEE.

[21] Loshchilov, I., & Hutter, F. (2016). Sgdr: Stochastic gradient descent with warm restarts. *arXiv preprint arXiv:1608.03983*.

[22] Zeiler, M. D., & Fergus, R. (2014, September). Visualizing and understanding convolutional networks. In *European conference on computer vision* (pp. 818-833). Springer, Cham.

[23] Kohli, D. (2015) Identifying and Classifying Slum Areas using Remote Sensing. Doctoral Dissertation: University of Twente Faculty of

Geo-Information Science and Earth Observation.

[24] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems* (pp. 1097-1105).