Locating Informal Urban Settlements

Bob Bell[1] , **Rajesh Veeraraghavan**[2]
[1]Georgetown University
[2]Georgetown University
{bob.bell, rajesh.veera}@georgetown.edu

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 high 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.

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]. The role of data and technology in making cities "smart" have gotten a lot of attention, but they are typically focused on the wealthy. India's government has proposed a nationwide program to build 100 smart cities that focus on such problems of the wealthy and the US has committed significant resources to supporting this initiative. At the same time, Rapid urbanization around the world and in India, have lead to the rise of mega cities, where the rural poor are increasingly migrating to cities in large numbers. Cites and governments are unable to cope with even basic information about where the poor live to adequately provide them with basic services. We don't know much about how the poor live as they migrate to the cities. The costs of doing surveys remain expensive, and it is particularly hard for resource starved countries in the global south to keep update the information about where the poor are. Yet, there is a promising research direction, where studies are using satellite imagery using night lights to find aggregate measures of where the poor live. There is a promising path set out by work done by Blumentstock that combine big data analysis with surveys, which point to a productive possibilities to estimates of poverty [Blumenstock, 2016]. This paper continues in that path, with satellite based work to understand how the poor are living.

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.

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"[2]. It is estimated that over the next few years India's infrastructure investment will be in the \$ 1.5 to \$ 2 trillion range. While smart cities have gotten a lot of recent attention in a city like Delhi, 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. The planned settlements in Delhi occupy only a quarter of Delhi's population, yet these plans of using data and technology reflect the global priorities where data amplifies existing inequalities [Toyama, 2011][Eubanks, 2018]. 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[Bhan and Jana, 2015]. There has been assessments based on qualitative studies that show that 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 [Bhan and Jana, 2015]. Yet, by its own admission, the Indian government and city governments have a very limited understanding of the spatial distribution of basic infrastructure in the poor parts of the Delhi and its socio-economic effects [Heller *et al.*, 2015].

There is a nascent quantitative literature on urban India that has drawn attention to the spatial dynamics of exclu-

[1]www.un.org/

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

sion [Sidhwani, 2015][Singh, 2014] [Singh and Vithayathil, 2012]. 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 [Heller *et al.*, 2015]. 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 [Heller *et al.*, 2015].

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 [Krizhevsky *et al.*, 2012]

The use of machine learning to detect informal settlements is an emerging area of research [Kuffer *et al.*, 2016]. 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. [Kuffer *et al.*, 2016]. Despite this, the accuracy of these studies hovered around 70-81% [Schmitt *et al.*, 2018].

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 [Stark, 2018].

World Bank conducted a study on Poverty from Space to predict poverty rates [Engstrom *et al.*, 2017]. Their study combined CNN's for car and roof detection but then incorporated additional feature engineering (PanTex, HOG, SIFT, etc) into a larger classifier.

2 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.

3 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]. We trained the model on a number of different ResNet architectures (ResNet-18/34/50/101/152) as well as VGG-19 [Simonyan and Zisserman, 2014], which was utilized in a similar study [Stark, 2018].

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 [Loshchilov and Hutter, 2016], 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 [Zeiler and Fergus, 2014]; (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. This cost approximately $2.70 on a GPU-enabled cloud computing platform.

3.1 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 (See 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, 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.

	Accuracy	
Architecture	*Pre-trained*	*Augmented Pre-trained*
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	82.23%	95.48%

Table 1: Accuracy based on ML architecture

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. 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.

3.2 Limitations

First, we subsample the "planned colonies" class in order to avoid imbalance. We do not compare our results with other data balancing strategies that avoid a reduction to training data (e.g., oversampling the rarer class). Future work will explore different sampling strategies that not only address imbalanced classes but also spatial sampling approaches that split based on regions within New Delhi.

Second, this paper's analysis is limited to two specific settlement types: planned colonies and JJCs. In Delhi, there are almost a dozen settlement types comprising residential, commercial, government, and industrial areas. Because of a lack of available data on the other settlement types, we restricted our analysis to the settlements in the city for which

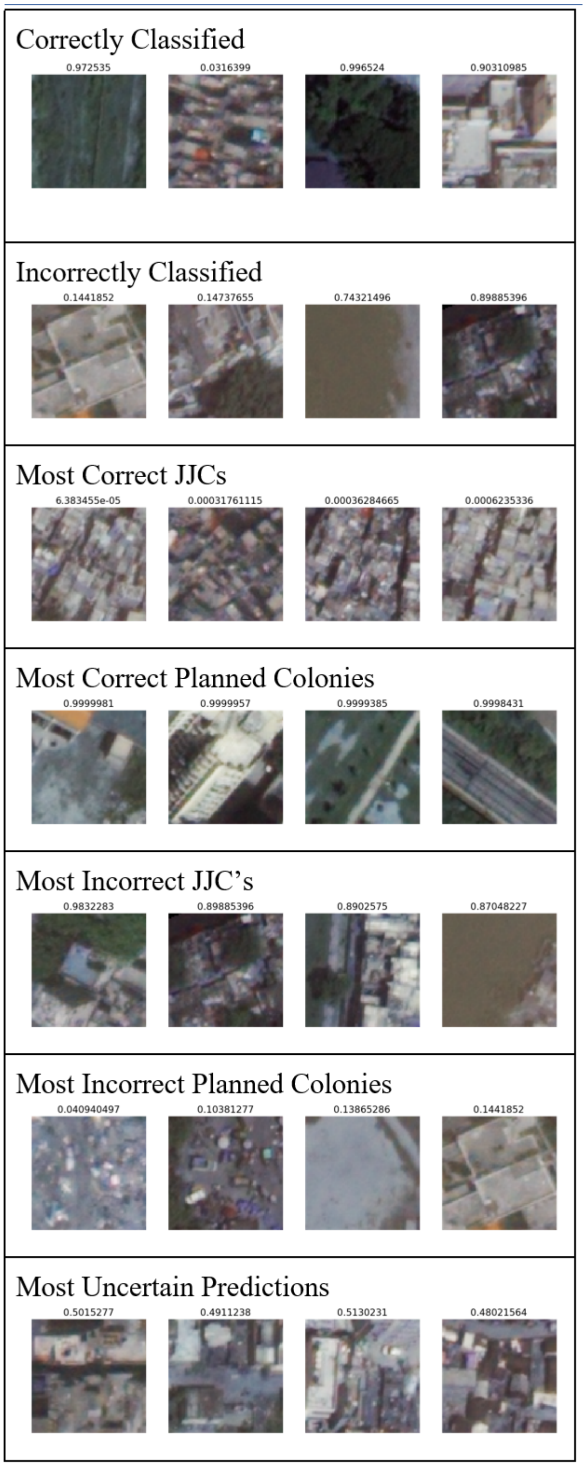

Figure 1: Illustration of the correct and incorrect classification examples from ResNet-34 pretrained model

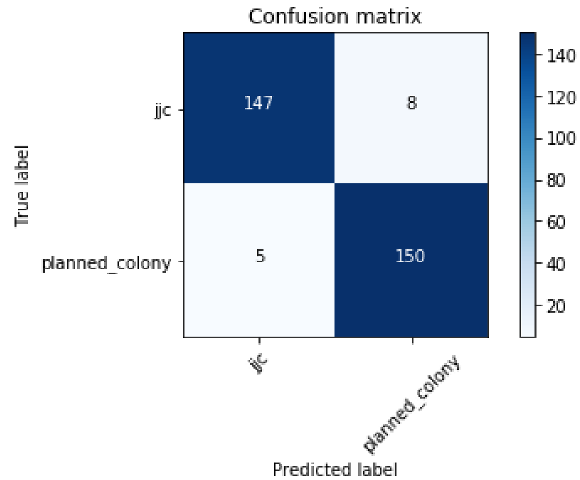

Figure 2: Confusion Matrix for Augmented ResNet-18

we had data. We are in the process of creating a comprehensive dataset of all settlements in Delhi, which will lead to a more fine-grained classification of the different settlements within the city.

3.3 Conclusions and Future Work

We built a classifier to automatically detect slums from planned colonies. 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.

The compute costs for training the classifier were minimal ($3), suggesting that as more low-cost and easy-to-use cloud services as well as pre-trained models 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, an approach that would be arguably cheaper than doing an actual slum survey. However, the cost of imagery is still prohibitively expensive.

In the near-term, it would be ideal to extend this classification work with high or 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

[Bhan and Jana, 2015] Gautam Bhan and Arindam Jana. Reading spatial inequality in urban india. *Economic and Political Weekly*, 50(22):49–54, 2015.

[Blumenstock, 2016] Joshua Evan Blumenstock. Fighting poverty with data. *Science*, 353(6301):753–754, 2016.

[Engstrom *et al.*, 2017] Ryan Engstrom, Jonathan Hersh, and David Newhouse. Poverty from space: using high-resolution satellite imagery for estimating economic well-being, 2017.

[Eubanks, 2018] Virginia Eubanks. *Automating inequality: How high-tech tools profile, police, and punish the poor*. St. Martin's Press, 2018.

[Heller *et al.*, 2015] Patrick Heller, Partha Mukhopadhyay, Subhadra Banda, and Shahana Sheikh. Exclusion, informality, and predation in the cities of delhi. *Centre for Policy Research*, 2015.

[Krizhevsky *et al.*, 2012] 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.

[Kuffer *et al.*, 2016] Monika Kuffer, Karin Pfeffer, and Richard Sliuzas. Slums from space—15 years of slum mapping using remote sensing. *Remote Sensing*, 8(6):455, 2016.

[Loshchilov and Hutter, 2016] Ilya Loshchilov and Frank Hutter. Sgdr: Stochastic gradient descent with warm restarts. *arXiv preprint arXiv:1608.03983*, 2016.

[Schmitt *et al.*, 2018] Andreas Schmitt, Tobias Sieg, Michael Wurm, and Hannes Taubenböck. Investigation on the 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, 2018.

[Sidhwani, 2015] Pranav Sidhwani. Spatial inequalities in big indian cities. *Economic & Political Weekly*, 50(22):55–62, 2015.

[Simonyan and Zisserman, 2014] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

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

[Singh, 2014] Gayatri Singh. *Freedom to move, barriers to stay: an examination of rural migrants' urban transition in contemporary India*. PhD thesis, PhD Dissertation, Department of Sociology, Brown University, 2014.

[Stark, 2018] Thomas Stark. Using deep convolutional neural networks for the identification of informal settlements to improve a sustainable development in urban environments. *Technische Universität München*, 2018.

[Toyama, 2011] Kentaro Toyama. Technology as amplifier in international development. In *Proceedings of the 2011 iConference*, pages 75–82. 2011.

[Zeiler and Fergus, 2014] Matthew D Zeiler and Rob Fergus. Visualizing and understanding convolutional networks. In *European conference on computer vision*, pages 818–833. Springer, 2014.